

**Modellierung posttranslationaler Regulation
innerhalb des Trisporsäure-Pheromonsystems
in Zygomyceten**

**Dissertation
zur Erlangung des akademischen Grades
„doctor rerum naturalium“ (Dr. rer. nat.)**

**vorgelegt dem Rat der Biologisch-Pharmazeutischen Fakultät
der Friedrich-Schiller-Universität Jena**

von Diplom-Bioinformatikerin Sabrina Ellenberger

geboren am 18.05.1987 in Arnstadt

Gutachter:

1. Prof. Dr. Johannes Wöstemeyer, Friedrich-Schiller-Universität Jena
2. Prof. Dr. Stefan Schuster, Friedrich-Schiller-Universität Jena
3. Prof. Dr. Sonja Prohaska, Universität Leipzig

Tag der öffentlichen Verteidigung: 27.10.2017

„Den Wissenschaftlern geht es wie den Chaoten.
Es ist alles da, man muss es nur suchen.“

Franz Kern

Danksagung

An dieser Stelle möchte ich mich bei allen bedanken, die mich bei meiner Promotion tatkräftig unterstützt haben.

Meinen besonderen Dank richte ich an meine beiden Doktorväter Johannes Wöstemeyer und Stefan Schuster. Danke Johannes, dass du dich auf das Experiment *Bioinformatik* eingelassen hast. Du hast mir gezeigt, dass man auch mal verrückt erscheinenden Ideen nachgehen sollte. Stefan danke ich für das entgegengebrachte Vertrauen und die Möglichkeit, bei der Entwicklung meiner Arbeit freie Hand zu haben.

Bei Anke Burmester möchte ich mich für ihre Bemühungen bedanken, mir die Organismen hinter den Sequenzdaten näher zu bringen.

Ein ganz besonderes Dankeschön geht an meine beiden Mitstreiterinnen Emilia Schulz und Lisa Siegmund für die interessanten Diskussionen und hilfreichen Anregungen während der letzten Jahre und besonders für den Zusammenhalt im kollektiven Wahnsinn, der dabei geholfen hat, ein Ventil für die angestaute Frustration beim Schreiben zu finden.

Außerdem möchte ich mich bei Bonny und Ivy bedanken, meinen beiden Maine Coon Mädchen, die mir immer wieder deutlich gemacht haben, dass es noch andere Dinge gibt, als Sequenzen und Proteinmodelle und sich in den entscheidenden Momenten der Verwirrung einfach auf dem Laptop niedergelassen haben.

Nicht zuletzt möchte ich meiner Mutter für ihre Unterstützung und Geduld danken, die sich gerade in der Schlussphase mit schlechter Laune, Geistesabwesenheit und Vernachlässigung arrangieren musste.

Zusammenfassung

In dem zur Großgruppe der Zygomyceten gehörenden Fusionsparasiten *Parasitella parasitica* (*Mucorales*) können sowohl die sexuelle Kommunikation, als auch die Kommunikation zwischen Wirt und Parasit in ein und demselben Organismus untersucht werden. Beide Interaktionen werden über das Trisporsäure-Pheromonsystem vermittelt. *Parasitella parasitica* infiziert viele verschiedene Zygomyceten. Besonders mit *Absidia glauca* als Wirtsorganismus ist dieser Infektionsprozess, genau wie die sexuelle Interaktion selbst, Kreuzungstyp-spezifisch. Das genetische System aus *Absidia glauca* und *Parasitella parasitica* ermöglicht sowohl die Erforschung der Interaktionen zwischen Wirt und Parasit, als auch des hier besonders häufig auftretenden horizontalen Gentransfers (HGT) in Eukaryoten. Um über das Labormodell hinaus den Umfang des Gentransfers zwischen Wirt und Parasit auch in evolutionären Zeiträumen zu untersuchen, wurden die Genome von *Absidia glauca* CBS 101.48 RVII-324 met⁻ und *Parasitella parasitica* CBS 412.66 NGI-315 ade⁻ sequenziert, annotiert und miteinander verglichen.

Dabei zeigte sich, dass Gene aus *Absidia glauca* eine unerwartet hohe Sequenzähnlichkeit zu Genen aus *Mucor*-artigen Zygomyceten aufweisen. Diese erhöhte Sequenzähnlichkeit beschränkt sich allerdings nicht nur auf *Parasitella parasitica*, was die Identifizierung von horizontalem Gentransfer zwischen Wirt und Parasit erschwert. Unter den putativen HGT-Genen finden sich vor allem transponierbare Elemente, wie Integrasen, Nukleasen und Transposasen. *Absidia glauca* und *Parasitella parasitica* besitzen eine für Pilze ungewöhnlich hohe Anzahl transponierbarer Elemente. Diese unterscheiden sich zum Teil zwischen Wirt und Parasit. Für einige dieser transponierbaren Elemente ist horizontaler Gentransfer dokumentiert. In mitochondrialer DNA gibt es ebenfalls transponierbare Elemente. Ein Vergleich der mitochondrialen DNA in Zygomyceten macht deutlich, dass die Mobilität genetischer Elemente mit Hilfe von Homing-Endonukleasen in Zygomyceten von Wirt-Parasit-Interaktionen unabhängig ist. Besonders ist hier die Ansammlung von Endonukleasen in den Introns der mitochondrialen DNA von *Parasitella parasitica*.

Ein zweites Kernthema dieser Arbeit ist die Modellierung der posttranslationalen Regulation von Enzymen innerhalb der Pheromonbiosynthese. Diese Enzyme finden sich in allen Zygomyceten. In *Parasitella parasitica* übernehmen sie weitere Funktionen und beeinflussen die differentielle Kommunikation bei sexuellen und parasitischen Interaktionen. 4-Dihydromethyltrisorat-Dehydrogenase (TSP1) ist eine NADP-abhängige Oxidoreduktase, die Trisporoide umsetzt. Homologiemodellierung und Protein-Protein-Docking wurden dazu verwendet, die Regulation der Enzymaktivität auf Proteinebene zu untersuchen. Das Enzym ist als Dimer aktiv. Bei *Parasitella parasitica* ist ein Muster bei der Enzymaktivität von TSP1-Dimeren erkennbar, das sexuelle und parasitische Interaktion voneinander unterscheidet. *Parasitella parasitica* profitiert von den Nährstoffen des Wirts, besonders von dessen Trisporsäure. *Parasitella parasitica* bildet erst sexuelle Sporen,

wenn mindestens einer der komplementären Kreuzungspartner auf einem Wirt wächst. Die vom Parasiten produzierte Trisporsäure reicht nicht aus, um diesen Effekt zu erzeugen. Die Analyse der TSP1-Dimere zeigt, dass der Parasit die Trisporsäure-Produktion des Wirts anregt, um das gesteigerte Pheromon-Niveau für die eigene sexuelle Entwicklung zu nutzen.

Die Arbeit wurde um die Suche nach alternativen Detektionsmethoden für horizontalen Gentransfer in Eukaryoten erweitert. HGT Calculator wurde als eine bewertungsbasierte Anwendung entwickelt, die vier unterschiedliche, sich ergänzende Ansätze zur Detektion von horizontalem Gentransfer zwischen prokaryotischen Endosymbionten und ihren protozoischen Wirten kombiniert. Das Programm wurde auf Gene aus Trypanosomatiden mit unterschiedlicher Lebensweise angewandt, um den Einfluss von Parasitismus auf die Häufigkeit von HGT-Ereignissen zu untersuchen.

Summary

In the fusion parasite *Parasitella parasitica* (*Mucorales*), which belongs to the fungal group of zygomycetes, sexual communication as well as host-parasite communication can be studied in the same organism. Both interactions are mediated by the trisporic acid pheromone system. *Parasitella parasitica* infects many different zygomycetes. Especially with *Absidia glauca* as host organism, the infection process is mating type dependent as sexual interactions are. The genetic system of *Absidia glauca* and *Parasitella parasitica* allows the investigation of host-parasite interactions and horizontal gene transfer (HGT) in eukaryotes, which occurs frequently in this genetic system. In order to investigate the extent of gene transfer between host and parasite beyond the laboratory model and on evolutionary time scales, the genomes of *Absidia glauca* CBS 101.48 RVII-324 met⁻ and *Parasitella parasitica* CBS 412.66 NGI-315 ade⁻ were sequenced, annotated and compared.

This comparison showed that genes from *Absidia glauca* yield an unexpected high sequence similarity to genes from *Mucor*-like zygomycetes. This high sequence similarity is not limited to *Parasitella parasitica* and hampers the identification of horizontal gene transfer between host and parasite. Among the putative HGT genes are at first transposable elements like integrases, nucleases and transposases. *Absidia glauca* and *Parasitella parasitica* have a high number of transposable elements quite unusual for fungi. They differ between host and parasite to some extent. For some of the transposable elements horizontal gene transfer is documented. Inside of mitochondrial DNA there are also transposable elements. A comparison of mitochondrial DNA in zygomycetes reveals that the mobility of genetic elements with the help of homing endonucleases is independent from host-parasite-interactions. A special feature is the accumulation of endonucleases inside of introns in the mtDNA of *Parasitella parasitica*.

The second main topic of this thesis is the modelling of post-translational regulation of enzymes of the pheromone biosynthesis. These enzymes occur in all zygomycetes. In *Parasitella parasitica*, they have additional functions and influence the differential communication in sexual and parasitic interactions. 4-Dihydromethyltrisporate dehydrogenase (TSP1) is a NADP-dependent oxidoreductase acting on trisporoids. Homology modelling and protein-protein docking were used to investigate the regulation of enzyme activity at the protein level. The enzyme is active as a dimer. In *Parasitella parasitica*, a pattern in enzyme activity of TSP1 dimers can be observed, which differentiates between sexual and parasitic interaction. *Parasitella parasitica* benefits from the nutrients of the host, especially the trisporic acid. *Parasitella parasitica* only forms sexual spores, if at least one mating partner is infecting a host. The level of trisporic acid produced by the fungus itself is too low to be effective. The analysis of TSP1 dimers shows that the parasite supports the trisporic acid production of the host to use the increased pheromone level for its own sexual development.

The thesis was extended by a search for alternative detection methods for horizontal gene transfer in eukaryotes. HGT Calculator was developed as a score-based application combining four different but complementary approaches for horizontal gene transfer detection between prokaryotic endosymbionts and their protozoan hosts. The tool was applied on genes from trypanosomatids with different life style to investigate the influence of parasitism on the frequency of HGT events.

Inhaltsverzeichnis

Zusammenfassung	5
Summary	7
1 Biologischer und bioinformatischer Hintergrund	12
1.1 <i>Absidia glauca</i> und <i>Parasitella parasitica</i> :	
Untersuchung der Kommunikation zwischen Wirt und Parasit	12
1.1.1 Das Infektionssystem	13
1.1.2 Trisporsäure, das Pheromon der Zygomyceten	13
1.1.3 Homothallie und Heterothallie	14
1.1.4 Asexuelle Entwicklung	15
1.1.5 Sexuelle Interaktion	16
1.1.6 Parasitische Interaktion	17
1.2 Horizontaler Gentransfer bei Eukaryoten	18
1.2.1 Transponierbare Elemente im Kerngenom	19
1.2.2 Mobile Elemente in mitochondrialer DNA	22
1.2.3 Endosymbiotischer und horizontaler Gentransfer	24
1.2.4 Etablierung endosymbiotischer Interaktionen	24
1.2.5 Detektion von horizontalem Gentransfer	27
1.3 Vorhersage von Proteinstrukturen	27
1.3.1 Das Grundproblem der Proteinfaltung	28
1.3.2 Empirische Strukturvorhersage	32
1.3.3 Homologie-basierte Strukturvorhersage	33
1.3.4 Strukturvorhersage mit künstlichen neuronalen Netzen	34
1.3.5 <i>De novo</i> Strukturvorhersage	35
2 Material und Methoden	38
2.1 Genomsequenzierung, -assemblierung und -annotierung	38
2.2 Mitochondriale DNA weiterer Zygomyceten	40
2.3 Identifizierung von TSP-Proteinen	41
2.4 Homologiemoellierung	42
2.5 Auswertung der modellierten Proteinstrukturen	46
2.6 Protein-Ligand-Docking mit AutoDock Vina	47
2.7 Protein-Protein-Docking mit ClusPro	54
2.8 Modellierung lokaler Proteindynamik mit PELE	55
2.9 Programmentwicklung	57
2.9.1 Detektion von horizontalem Gentransfer: HGT Calculator	57

2.9.2	Genvorhersage für <i>Absidia glauca</i> und <i>Parasitella parasitica</i> : ZygoViewer	69
3	Ergebnisse	72
3.1	Vergleich der Genome von Wirt und Parasit	72
3.1.1	Das Genom von <i>Absidia glauca</i>	72
3.1.2	Das Genom von <i>Parasitella parasitica</i>	73
3.1.3	Sequenzähnlichkeit der für <i>Absidia glauca</i> vorhergesagten Gene .	73
3.1.4	Analyse der mitochondrialen DNA von Zygomyceten	74
3.1.5	Transponierbare Elemente im Kerngenom	86
3.2	Strukturmodellierung von TSP1, TSP2 und TSP3	90
3.2.1	4-Dihydromethyltrisporat-Dehydrogenase (TSP1)	90
3.2.2	4-Dihydrotrisporin-Dehydrogenase (TSP2)	109
3.2.3	Carotin-Oxygenase (TSP3)	121
3.3	Analyse der Ligandenbindung in TSP1- und TSP2-Proteinen	128
3.3.1	Ligandenbindung bei TSP1	128
3.3.2	Substratspezifität für Trisporoide der B- und E-Serie in TSP1 . . .	133
3.3.3	Ligandenbindung bei TSP1 in <i>Parasitella parasitica</i>	134
3.3.4	Ligandenbindung bei TSP2	134
3.3.5	Substratspezifität für Trisporoide der B- und E-Serie in TSP2 . . .	138
3.3.6	Ligandenbindung bei TSP2 in <i>Parasitella parasitica</i>	138
3.3.7	Gemeinsamkeiten in der Ligandenbindung bei TSP1 und TSP2 . .	139
3.4	Untersuchung der Bildung von Homo- und Heterodimeren	141
3.4.1	Modellierung von TSP1-Dimeren in <i>Absidia glauca</i>	142
3.4.2	Modellierung von TSP1-Dimeren in <i>Parasitella parasitica</i>	144
3.4.3	Modellierung von TSP2-Dimeren in <i>Parasitella parasitica</i>	147
3.4.4	Modellierung von TSP1–TSP2-Dimeren in <i>Parasitella parasitica</i>	148
3.4.5	Modellierung von TSP1-Dimeren zwischen <i>Absidia glauca</i> und <i>Parasitella parasitica</i>	150
3.5	<i>In silico</i> Mutanten von TSP1 aus <i>Parasitella parasitica</i> : Aktivierung von TSP1-Homodimeren	152
3.6	Detektion von horizontalem Gentransfer zwischen Bakterien und Protozoen	155
3.6.1	Gezielte Identifizierung von HGT in Isocitrat-Dehydrogenasen . .	155
3.6.2	Gezielte Identifizierung von HGT in <i>Trypanosoma rangeli</i>	159
4	Diskussion	162
4.1	Vergleich der Genome von Wirt und Parasit	162
4.1.1	Kerngenom	162
4.1.2	Mitochondriale DNA	163

4.2	Posttranslationale Regulation innerhalb der Trisporsäure-Synthese	166
4.2.1	Posttranslationale Regulation von TSP1	166
4.2.2	Posttranslationale Regulation von TSP2	168
4.2.3	Interaktion zwischen TSP1 und TSP2	170
4.3	Detektion von horizontalem Gentransfer zwischen Protozoen und Bakterien	170
4.3.1	Identifizierung von horizontalem Gentransfer	171
4.3.2	Vorteile gezielter Identifizierung von horizontalem Gentransfer . .	172
5	Literatur	176
6	Anhang	191
6.1	Zusätzliches Material	191
6.2	Abkürzungsverzeichnis	192
6.3	Abbildungsverzeichnis	193
6.4	Wissenschaftliche Beiträge	196
6.4.1	Veröffentlichung von Genomen und Software	196
6.4.2	Publikationen	196
6.4.3	Poster	198
6.5	Ehrenwörtliche Erklärung	200

1 Biologischer und bioinformatischer Hintergrund

Das in dieser Arbeit untersuchte genetische System besteht aus den beiden Pilzen *Parasitella parasitica* und *Absidia glauca*. Der Fusionsparasit *Parasitella parasitica* infiziert viele verschiedene Pilze (Zygomyceten) (Burgeff, 1924). Neben heterothallischen Zygomyceten (siehe Abschnitt 1.3), wie zum Beispiel *Absidia glauca* oder *Mucor mucedo*, werden auch homothallische Zygomyceten (siehe Abschnitt 1.2), wie *Mycotypha africana* oder *Zygorhynchus moelleri*, von *Parasitella parasitica* infiziert. Zu den Zygomyceten, bei denen keine Infektion beobachtet wurde, zählen unter anderem *Phycomyces blakesleeana* und *Syncephalastrum racemosum*.

1.1 *Absidia glauca* und *Parasitella parasitica*: Untersuchung der Kommunikation zwischen Wirt und Parasit

Bei den untersuchten Zygomyceten-Stämmen *Absidia glauca* CBS 101.48 RVII-324 met⁻ und *Parasitella parasitica* CBS 412.66 NGI-315 ade⁻ handelt es sich um auxotrophe Mutanten. Diese produzieren aufgrund von Mutationen in spezifischen Genen bestimmte Aminosäuren oder Nukleotide nicht selbst. Für ein optimales Wachstum sind sie auf eine Nährstoffquelle angewiesen, die diese fehlenden Substanzen bereitstellt. *Absidia glauca* CBS 101.48 RVII-324 met⁻ ist eine mit Röntgenstrahlung erzeugte Methionin-Auxotrophe des *Absidia glauca* CBS 101.48 (–)-Kreuzungstyps (Wöstemeyer et al., 1990; Kellner et al., 1993). *Parasitella parasitica* CBS 412.66 NGI-315 ade⁻ ist eine nach Behandlung mit Nitrosoguanidin entstandene Adenin-Auxotrophe des *Parasitella parasitica* CBS 412.66 (+)-Kreuzungstyps (Burmester et al., 2013).

Absidia glauca ist nicht so eng mit *Parasitella parasitica* verwandt, wie andere potentielle Wirte, zum Beispiel *Mucor circinelloides*. Die verwandtschaftliche Nähe einiger Wirte zu *Parasitella parasitica* lässt sich anhand der Sequenzidentität ihrer Aktin-Gene veranschaulichen. Diese Sequenzähnlichkeit liegt auf Nukleotidebene bei *Absidia glauca* und *Parasitella parasitica* bei 83 % und entspricht damit der durchschnittlichen Sequenzähnlichkeit von Aktin in Zygomyceten. Zwischen *Mucor circinelloides* und *Parasitella parasitica* liegt sie hingegen bei 95 %. Beide Pilze gehören zur Gruppe der *Mucor*-artigen Zygomyceten. *Absidia glauca* gehört zur Familie der Cunninghamellaceae. Vergleicht man Pilze innerhalb dieser beiden Gruppen, wie zum Beispiel *Mucor ambigua* und *Mucor circinelloides* (96 %) oder *Absidia repens* und *Absidia glauca* (93 %) sind die Sequenzidentitäten entsprechend höher. Die Sequenzidentität von Aktin-Genen aus *Parasitella parasitica* und *Phycomyces blakesleeana*, der nicht als Wirt dient, beträgt 81 %. Die phylogenetische Verwandtschaft zwischen dem Parasiten *Parasitella parasitica* und weiteren potentiellen Wirten wurde in einer Veröffentlichung von Hoffmann et al. (2013) beschrieben. In dieser Studie wurden auch die beiden Wildtyp-Stämme von *Ab-*

sidia glauca und *Parasitella parasitica* berücksichtigt. Mit Hilfe der großen und kleinen Untereinheiten ribosomaler DNA, ITS-Regionen (internal transcribed spacer), Aktin und des Translationselongationsfaktors 1-alpha wurde hier die Familienstruktur der *Mucorales* untersucht.

1.1.1 Das Infektionssystem

In Zygomyceten werden sowohl sexuelle, als auch parasitische Interaktionen über dasselbe Kommunikationssystem vermittelt. Dieses Pheromonsystem basiert auf dem Apocarotinoid Trisporsäure und dessen biosynthetischen Vorstufen. In *Parasitella parasitica* können sexuelle und parasitische Kommunikation zur gleichen Zeit in einem Organismus beobachtet und untersucht werden (Schultze et al., 2005). Während des asexuellen Entwicklungszyklus entstehen bei *Parasitella parasitica* Sporangiosporen. *Parasitella parasitica* ist heterothallisch (siehe Abschnitt 1.1.3) und entwickelt zusammen mit einem komplementären Kreuzungspartner sexuelle Sporen (Zygosporen, Abbildung 1.6 (d)). *Parasitella parasitica* ist ein Fusionsparasit (siehe Abschnitt 1.1.6), der verschiedene Zygomyceten infiziert. Ein schon früh erkanntes und inzwischen gut untersuchtes System zur Erforschung der Wechselwirkungen zwischen Wirt und Parasit verwendet den Zygomyceten *Absidia glauca* als Wirtsorganismus (Burgeff, 1924). Dieses System findet auch bei der Untersuchung von horizontalem Gentransfer in Eukaryoten Verwendung (Burmesster et al., 2013). Der Infektionsprozess ist bei *Absidia glauca* ebenfalls Kreuzungstypspezifisch. Das heißt, ein (+)-Kreuzungstyp von *Parasitella parasitica* infiziert nur *Absidia glauca*-Stämme vom (–)-Kreuzungstyp und umgekehrt.

1.1.2 Trisporsäure, das Pheromon der Zygomyceten

Jeder Zygomycet produziert und verarbeitet eine Mischung aus unterschiedlichen Trisporoid-Derivaten, die verschiedenen Serien zugeordnet werden (Miller und Sutter, 1984). Diese werden zu veränderlichen Anteilen zur Herstellung des Pheromons Trisporsäure verwendet (Werkman, 1976).

Diese Arbeit konzentriert sich hauptsächlich auf die Trisporoid-Varianten der B- und E-Serie. Es gibt eine Präferenz für Trisporoide der B-Serie in *Mucor mucedo* und für Trisporoide der E-Serie in *Phycomyces blakesleeanus* (Schimek und Wöstemeyer, 2009; Sutter et al., 1989).

Trisporoide unterscheiden sich anhand ihrer Substituenten an den C1- und C4-Atomen. Die Substituenten an den C2- und C3-Atomen des Rings und am C13-Atom am Ende der Seitenkette des Moleküls charakterisieren die unterschiedlichen Derivate (Bu'Lock et al., 1972; Schimek und Wöstemeyer, 2009; Sutter et al., 1989). Trisporoide der B-Serie haben am C13-Atom eine Ketogruppe, Trisporoide der E-Serie hingegen eine Hydroxylgruppe. Zusätzlich haben Trisporoide der E-Serie eine weitere Hydroxylgruppe am C3-Atom.

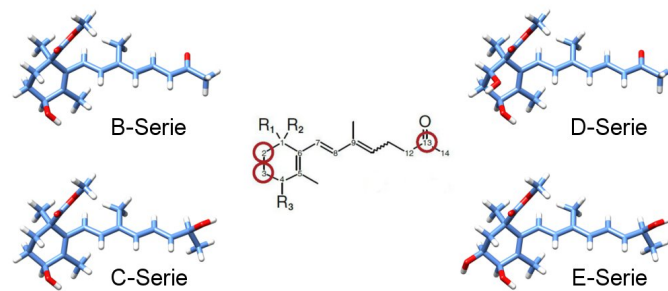


Abbildung 1.1: Strukturen der am häufigsten auftretenden Trisporoid-Typen.

Das Substrat der 4-Dihydromethyltrisorat-Dehydrogenase (TSP1) ist 4-Dihydromethyltrisorat:
 $R_1 = CH_3, R_2 = COOCH_3, R_3 = OH$

4-Dihydrotrisorin ist das Substrat der 4-Dihydrotrisorin-Dehydrogenase (TSP2). Dieses Trisporoid wurde dazu benutzt, die Substratspezifität von TSP1 zu untersuchen:

$R_1 = CH_3, R_2 = CH_3, R_3 = OH$

Abbildung 1.1 zeigt die Unterschiede zwischen den einzelnen Trisporoid-Varianten.

Die Tertiärstrukturen der TSP1- und TSP2-Proteine aus unterschiedlichen Zygomyceten liefern Hinweise auf die Substratspezifität der einzelnen Pilze. Da die dreidimensionalen Proteinstrukturen für TSP-Proteine nicht bekannt sind, wurden diese mit Homologiemodellierung erzeugt (siehe Abschnitt 2.4) und mit Protein-Ligand-Docking (siehe Abschnitt 2.6) auf ihre Substratspezifität untersucht.

1.1.3 Homothallie und Heterothallie

Bei homothallischen Zygomyceten entstehen die Gametangien während der sexuellen Entwicklung aus den Hyphen desselben Myzels (Abbildung 1.2). Es wird kein Kreuzungspartner benötigt. Stattdessen kommt es zu einer Spezialisierung der Zygophoren aus (+)- und (–)-artigen Hyphenenden (Abbildung 1.2 (a)). Diese Hyphenenden schwellen an, wodurch sich Progametangien bilden (Abbildung 1.2 (b)). Der Einzug von Septen führt zur Entstehung zweier Gametangien. Diese fusionieren (Abbildung 1.2 (c)) und bilden ein Zygosporangium (Abbildung 1.2 (d)). Das Zygosporangium differenziert sich zur reifen Zygospore aus (Abbildung 1.2 (e)). Diese ist pigmentiert. In dieser Arbeit wurden drei homothallische Zygomyceten berücksichtigt, *Mucor heterogamus*, *Mycotypha africana* und *Syncephalastrum racemosum*.



Abbildung 1.2: Generelles Schema der Entstehung von Zygosporen bei homothallischen Zygomyceten. Die (+)- und (–)-artigen Hyphenenden eines Myzels wachsen aufeinander zu (a), treffen sich und schwellen an. Es bilden sich Progametangien (b). Nach dem Einzug von Septen fusionieren die auf diese Weise entstandenen Gametangien (c) und es bildet sich ein Zygosporangium (d). Dieses reift zur fertigen Zygospore heran (e). (Abbildung nach Schulz und Wetzel (2016))

Bei heterothallischen Zygomyceten treten mehrere genetisch unterschiedliche Kreuzungstypen auf. Erst nach dem Aufeinandertreffen zweier komplementärer Kreuzungstypen, (+)- und (–)-Kreuzungstyp (Blakeslee, 1904), kommt es zur Differenzierung sexueller Strukturen (Abbildung 1.3). Die Zygophoren komplementärer Kreuzungspartner wachsen aufeinander zu (Abbildung 1.3 (a)). Sie schwellen an und bilden Progametangien (Abbildung 1.3 (b)). Diese werden mit Septen abgegrenzt und es bilden sich zwei Gametangien, die anschließend fusionieren (Abbildung 1.3 (c)). Das entstandene Zygosporangium (Abbildung 1.3 (d)) differenziert sich zur reifen pigmentierten Zygospore aus (Abbildung 1.3 (e)).



Abbildung 1.3: Generelles Schema der Entstehung von Zygosporen bei heterothallischen Zygomyceten wie *Mucor mucedo*. Die Zygophoren zweier komplementärer Kreuzungspartner wachsen aufeinander zu (a). Die Hyphenenden schwellen an der Kontaktstelle an und es bilden sich Progametangien (b). Diese werden mit Septen abgegrenzt und es entstehen zwei Gametangien, die anschließend fusionieren (c). Das gebildete Zygosporangium (d) differenziert sich zur reifen Zygospore aus (e). (Abbildung nach Schimek und Wöstemeyer (2009))

1.1.4 Asexuelle Entwicklung

Bei der asexuellen Entwicklung werden Mitosporen in Sporangien gebildet. Abbildung 1.4 zeigt die Sporangienbildung bei *Absidia glauca*. Hier werden die birnenförmigen Sporangien in Wirteln auf Lufthyphen angeordnet (Webster, 1983). Es kommen auch kugelförmige Sporangien vor. Sporangien entstehen entweder einzeln an der Spitze einer geraden Sporangiphore oder zu mehreren an einer verzweigten Sporangiphore.

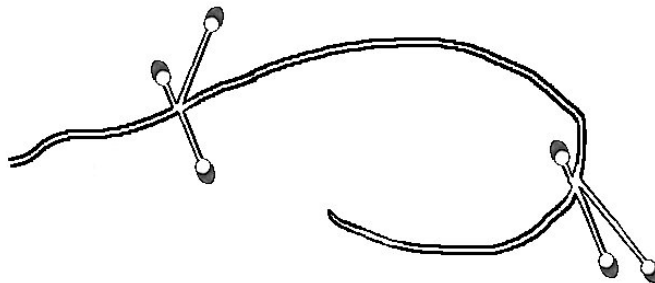


Abbildung 1.4: Asexuelle Entwicklung bei Zygomyceten. Bei *Absidia glauca* bilden sich Sporangien in Wirteln aus. (Abbildung nach Webster (1983))

1.1.5 Sexuelle Interaktion

Die Biosynthese des in Zygomyceten vorkommenden Pheromons Trisporsäure beginnt mit der oxidativen Spaltung von β -Carotin mit Hilfe einer Carotin-Oxygenase (TSP3, siehe Abschnitt 3.2.3) (van den Ende et al., 1972; Burmester et al., 2007; Schimek und Wöstemeyer, 2009). Nach mehreren noch nicht genau bekannten Reaktionen endet der Kreuzungstyp-unspezifische Syntheseweg bei 4-Dihydrotrisorin. Ab diesem Punkt sind die beiden Kreuzungspartner (+/–) von einem Austausch der Synthese-Zwischenprodukte abhängig. Allein ist keiner der beiden Kreuzungspartner in der Lage, ausreichende Mengen an Trisporsäure herzustellen.

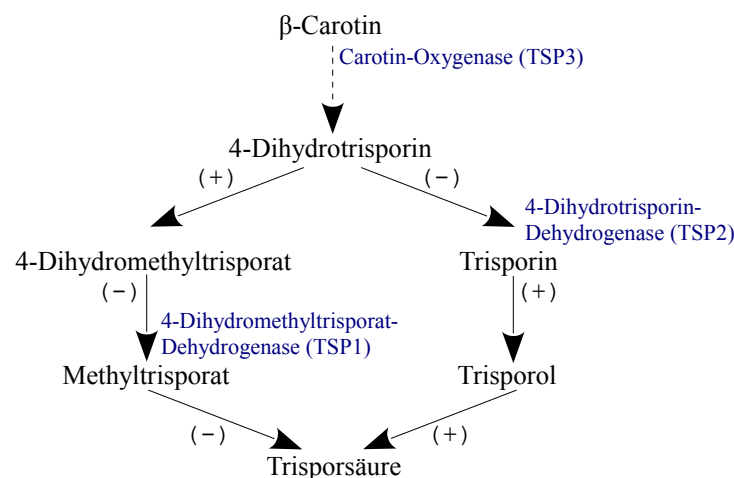


Abbildung 1.5: Kooperative Biosynthese des Pheromons Trisporsäure in Zygomyceten.

Die Kreuzungspartner (+) und (–) sind nicht in der Lage, Trisporsäure ohne einen Austausch von Zwischenprodukten allein in ausreichenden Mengen zu synthetisieren.

(Abbildung nach Werkman und van den Ende (1974))

Im (–)-Kreuzungstyp oxidiert eine NADP-abhängige Dehydrogenase, die 4-Dihydrotrisorin-Dehydrogenase (TSP2, siehe Abschnitt 3.2.2) (Wetzel et al., 2009) 4-Dihydrotrisorin zu Trisorin. Eine andere NADP-abhängige Dehydrogenase, die 4-Dihydromethyltrisorat-Dehydrogenase (TSP1, siehe Abschnitt 3.2.1) (Czempinski et al., 1996) wandelt 4-Dihydromethyltrisorat zu Methyltrisorat um. Im (+)-Kreuzungstyp wird 4-Dihydrotrisorin zu 4-Dihydromethyltrisorat oxidiert und Trisorin zu Trisorol umgewandelt. Die Enzyme, die diese Schritte ermöglichen, wurden noch nicht identifiziert. Abbildung 1.5 (Werkman und van den Ende, 1974; Werkman, 1976; Schimek und Wöstemeyer, 2006) zeigt den Ablauf der kooperativen Trisoroidsynthese. Dieses kooperative Synthese-Modell wurde von einem bioinformatischen Ansatz als die beste Möglichkeit der Trisporsäure-Produktion bestätigt (Werner et al., 2012).

Die Zygosporienentwicklung bei *Parasitella parasitica* ist typisch für *Mucor*-artige Zygomyceten (siehe Abbildung 1.3). Zygophoren-Induktion und gerichtetes Wachstum der Hyphen führen dazu, dass die Lufthyphen eines (+)-Kreuzungstyps und die Lufthyphen eines (–)-Kreuzungstyps zusammentreffen.

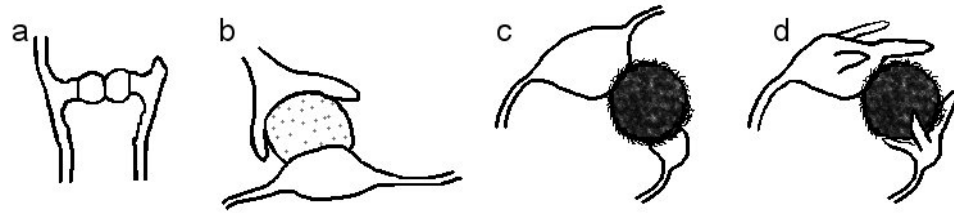


Abbildung 1.6: Sexuelle Entwicklung bei *Parasitella parasitica*.

Die Lufthyphen der Kreuzungspartner (+) und (–) wachsen aufeinander zu, berühren sich und bilden über mehrere Vorstufen Zygosporen aus, die wie die Zygosporen von *Mucor mucedo* aussehen. (Abbildung nach Burgeff (1924))

Es kommt zur Entstehung von Progametangien und Septenbildung (Abbildung 1.6 (a)). Es entstehen oft Finger-förmige Ausbildungen, die an die Krall-Hyphen bei *Absidia glauca* erinnern. Die Gametangien der beiden Kreuzungspartner fusionieren (Abbildung 1.6 (b)) und bilden eine Zygospore (Abbildung 1.6 (c)). Abbildung 1.6 (d) zeigt die reife Zygospore. Diese kommt bei *Parasitella parasitica* so gut wie nie zur Keimung.

1.1.6 Parasitische Interaktion

Der parasitische Lebensstil von *Parasitella parasitica* wird von der Fusion spezialisierter, räumlich beschränkter Strukturen an den Hyphenenden mit den Wirtshyphen begleitet. Der Fusionsparasitismus ist mit dem Transport von Organellen vom Parasiten zum Wirt verknüpft (Abbildung 1.7) und verursacht regelmäßig die Bildung genetischer Rekombinanten zwischen Wirt und Parasit (Burgeff, 1924; Burmester et al., 2013; Kellner et al., 1993; Wöstemeyer et al., 1995). Nachdem die Hyphen von Wirt und Parasit miteinander in Kontakt gekommen sind (Abbildung 1.7 (a)), kommt es zum Anschwellen der Hyphe von *Parasitella parasitica*. Diese Struktur wird als primäre sikyotische Zelle bezeichnet (Abbildung 1.7 (b,c)). Zellkerne aus *Parasitella parasitica* wandern in diesen Bereich. Der Abschnitt wird anschließend von einem Septum begrenzt (Abbildung 1.7 (b)). Die primäre sikyotische Zelle fusioniert mit der Wirtshyphe, was zur Plasmafusion führt. Es kommt zum Zusammentreffen und möglicher Rekombination der Zellkerne aus Wirt und Parasit. Die sekundäre sikyotische Zelle wird gebildet. Diese nennt man auch Schröpfkopf (Abbildung 1.7 (c, d)). Dieser differenziert sich zur reifen Sikyospore (Abbildung 1.7 (e)) aus, die von einer Gallenstruktur umschlossen wird.

Bei der Bildung des cytoplasmatischen Kontinuums schädigt der Parasit den Wirt nur wenig und profitiert von den Ressourcen, die der Wirt besitzt. Zur Erkennung eines geeigneten Wirts wird das Trisporsäure-Pheromonsystem benutzt. Der Prozess ist, genau wie die sexuelle Interaktion, Kreuzungstyp-spezifisch. Ein (+)-Kreuzungstyp von *Parasitella parasitica* infiziert *Absidia glauca*-Stämme vom (–)-Kreuzungstyp und ein (–)-Kreuzungstyp von *Parasitella parasitica* infiziert *Absidia glauca*-Stämme vom (+)-Kreuzungstyp.

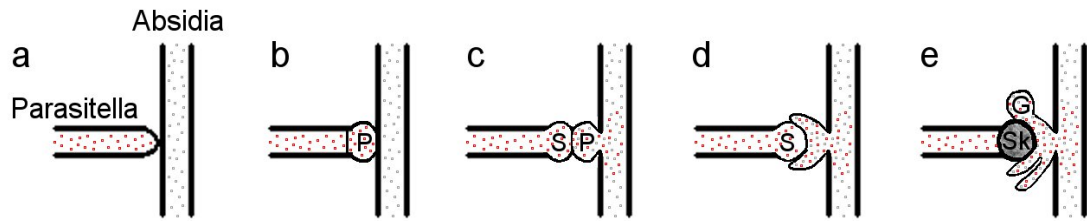


Abbildung 1.7: Schematische Darstellung des Fusionsparasitismus von *Parasitella parasitica*. Die Hyphe, die von oben nach unten verläuft, gehört dem Wirt *Absidia glauca* (graue Kreise). Die Hyphe, die sich der Wirtshyphe von links nähert, gehört dem Parasiten *Parasitella parasitica* (braune Kreise). Kreise symbolisieren die Zellkerne der beiden Pilze. Diese vermischen sich nach der Fusion. P: primäre sikyotische Zelle, S: sekundäre sikyotische Zelle, Schröpfkopf, G: Galle, Sk: Sikyospore (Abbildung nach Burgeff (1924))

1.2 Horizontaler Gentransfer bei Eukaryoten

Horizontaler Gentransfer (HGT) bezeichnet den Austausch genetischen Materials zwischen verschiedenen Spezies, die nicht über eine direkte Verwandtschaft miteinander in Beziehung stehen, das heißt, Nachkommen eines Vorfahrens sind. In Prokaryoten spielt horizontaler Gentransfer eine wichtige Rolle (Koonin et al., 2001; Popa und Dagan, 2011; Polz et al., 2013). Dieser Mechanismus fördert die Herausbildung selektiver Vorteile auf Seiten des Empfängers der fremden DNA und die Anpassung an neue ökologische Bedingungen (Marri et al., 2007). Beispiele hierfür sind die Aufnahme eines Antibiotikaresistenzgens, sowie die Ausprägung pathogener oder metabolischer Eigenschaften (Juhász et al., 2009). Die Mehrheit der Gentransfere hat allerdings keinen direkten Einfluss auf den Empfänger (Gogarten und Townsend, 2005). Horizontaler Gentransfer wird inzwischen auch immer häufiger in Eukaryoten identifiziert. In Protisten, die phagotroph leben, tritt horizontaler Gentransfer mit einer ähnlichen Häufigkeit auf, wie bei Prokaryoten (Andersson, 2005). Horizontaler Gentransfer findet auch in vielzelligen Organismen statt (Andersson, 2005). Pilze interagieren mit Bakterien, Pflanzen und anderen Pilzen und nehmen auf diesem Weg fremdes genetisches Material auf. Horizontaler Gentransfer existiert zwischen Prokaryoten und Pilzen (Prolin-Racemase, Phenazin F-homologes Protein) (Fitzpatrick et al., 2008). Andere HGT-Ereignisse führten zum Austausch von Genen zwischen Pilzen und Pflanzen (Alkohol-Dehydrogenase, Fucose-Permease, Membrantransporter, Zinkfinger-Protein) (Richards et al., 2009). Es gibt verschiedene Mechanismen, die horizontalen Gentransfer vermitteln. In Pilzen ist horizontaler Gentransfer für Plasmide, Mycoviren, Introns, transponierbare Elemente, Gene aus dem Kerngenom, Gencluster und ganze Chromosomen dokumentiert (Rosewich und Kistler, 2000).

1.2.1 Transponierbare Elemente im Kerngenom

Transposons sind genetische Elemente, die sich innerhalb eines Genoms vervielfältigen und ihre Position verändern. Sie kommen bei Archaeen, Bakterien und Eukaryoten vor und wurden als erstes von McClintock (1950) in Mais entdeckt.

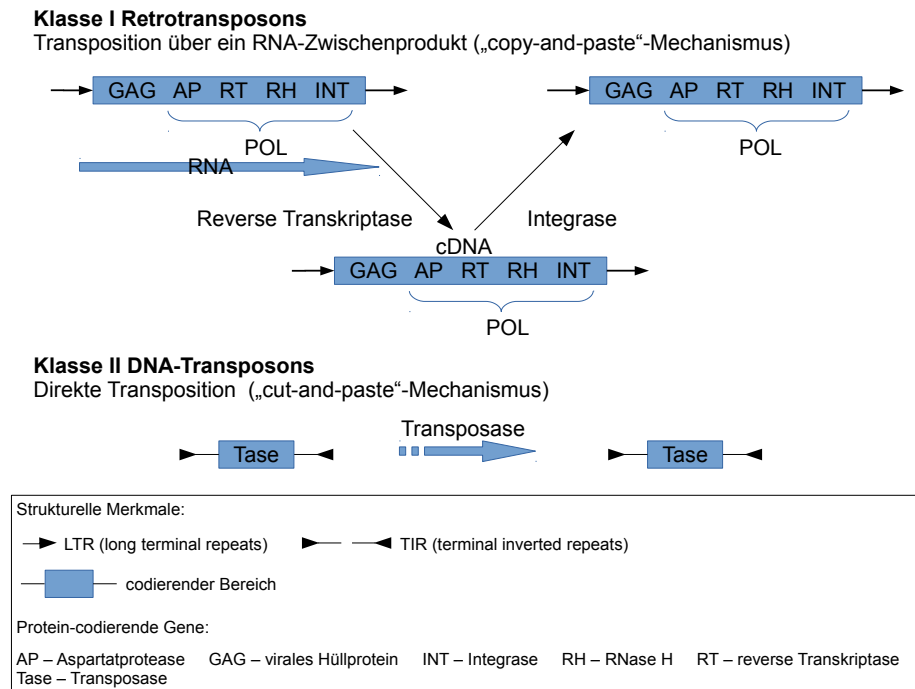


Abbildung 1.8: Mechanismus der Transposition von Transposons aus Klasse I und II in Eukaryoten. Transponierbare Elemente der Klasse I erzeugen in jedem Replikationszyklus eine neue Kopie von sich, die an einer anderen Stelle ins Genom integriert wird. Transponierbare Elemente der Klasse II verlassen ihre alte Position und reintegrieren an einer neuen Stelle im Genom. (Abbildung nach Fávoro et al. (2005))

Das erste System zur Klassifizierung von transponierbaren Elementen stammt von Finnegan (1989) und unterscheidet zwei Klassen (Abbildung 1.8). Bei Klasse I Retrotransposons vermittelt ein „Copy-and-Paste“-Mechanismus die Transposition. Dabei entsteht ein RNA-Zwischenprodukt als genomische Kopie. Diese wird von einer reversen Transkriptase, die in dem transponierbaren Element codiert ist, in DNA umgewandelt. Jeder vollendete Replikationszyklus erzeugt eine neue Kopie, die mit Hilfe einer Integrase an einer anderen Stelle ins Genom eingefügt wird.

Die einzelnen LTR-Retrotransposons (LTR: long terminal repeat) unterscheiden sich in der Anordnung ihrer Protein-codierenden Domänen (Aspartatprotease, Integrase, RNase H, reverse Transkriptase) innerhalb der POL-Region und beinhalten ein GAG-Gen, das für ein virales Hüllprotein codiert. Die Länge von LTR-Retrotransposons variiert zwischen ein paar hundert Basenpaaren und bis zu 25 kb.

Um Retroviren in der Klassifizierung transponierbarer Elemente zu berücksichtigen, die nach Inaktivierung und Deletion der Domänen für extrazelluläre Mobilität in LTR-

Retrotransposons transformiert wurden, wurden diese als sogenannte endogene Retroviren (ERV) in die Liste der LTR aufgenommen.

Eine andere neu beschriebene Gruppe sind die *DIRS*-ähnlichen Elemente (*DIRS*: *Dic-tyostelium* intermediate repeat sequence). Diese ähneln LTRs und beinhalten eine Tyrosin-Rekombinase statt einer Integrase.

LINEs (long interspersed nuclear elements) besitzen keine LTRs und haben eine Länge von mehreren kb. Autonome LINEs codieren für die Transposition mindestens eine reverse Transkriptase und eine Endonuklease in ihrer POL-Region.

Bei Klasse II DNA-Transposons vermittelt ein „Cut-and-Paste“-Mechanismus die Transposition (Abbildung 1.8). Die Anzahl dieser transponierbaren Elemente erhöht sich, wenn die Transposition während der Replikation von Chromosomen stattfindet. Dabei muss das Element an einer Stelle ausgeschnitten werden, die bereits repliziert wurde und an einer Stelle integrieren, die die Replikationsgabel noch nicht passiert hat.

TIR-DNA-Transposons (TIR: terminal inverted repeat) bestehen meist nur aus einer terminalen invertierten Wiederholung und einem Gen für eine Transposase, die die Transposition vermittelt. Diese erkennt TIRs und durchtrennt beide DNA-Stränge an jedem Ende.

Helitron-Elemente aus der RC-Familie benutzen einen der „rolling-circle“-Replikation (RC) ähnlichen Mechanismus, um sich zu replizieren (Thomas und Pritham, 2015). Dabei wird der DNA-Strang nur einmal zerschnitten. Als Matrize fungiert ein einzelsträngiges DNA-Zwischenprodukt. Die DNA-Polymerase folgt dem Minus-Strang wie einem rollenden Ring (rolling circle). Dabei wird der ursprüngliche Plus-Strang verdrängt. Am entstehenden Einzelstrang wird der komplementäre Strang zunächst als Okazaki-Fragmente synthetisiert und anschließend zu einem zusammenhängenden Strang ligiert.

Weitere neu entdeckte transponierbare Elemente, die einen „Copy-and-Paste“-Mechanismus besitzen, jedoch ohne ein RNA-Zwischenprodukt arbeiten, und hochgradig reduzierte nicht-autonome transponierbare Elemente, sogenannte MITEs (miniature inverted repeat transposable elements) haben eine Erweiterung der ursprünglichen Klassifizierung nach Finnegan (1989) notwendig gemacht. Wicker et al. (2007) behalten die zwei Klassen bei und erweitern sie um zusätzliche Klassifizierungsebenen. Tabelle 1.1 zeigt die Klassifizierung transponierbarer Elemente nach Wicker et al. (2007).

Die Klassifizierung nach Wicker et al. (2007) beinhaltet Klassen, Unterklassen, Ordnungen, Superfamilien, Familien und Unterfamilien. Die Klassen unterteilen transponierbare Elemente danach, ob sie ein RNA-Zwischenprodukt bei der Transposition verwenden. Unterklassen unterscheiden Elemente, die sich selbst vor einer Insertion kopieren, von solchen, die die Donorsequenz verlassen und sich an anderer Stelle einfügen. Ordnungen erfassen wesentliche Unterschiede bei den Mechanismen für Insertion. Superfamilien einer Ordnung teilen eine Replikationsstrategie und unterscheiden sich zum Beispiel in der Struktur Protein-codierender oder nicht-codierender Bereiche.

Klasse I (Retrotransposons)		
LTR (long terminal repeat)	<i>Copia</i>	Transposition <i>via</i> „Copy-and-Paste“-Mechanismus
	<i>Gypsy</i>	
	<i>Bel-Pao</i>	
	<i>Retrovirus</i>	
	<i>ERV</i> (endogener Retrovirus)	
DIRS (<i>Dictyostelium</i> intermediate repeat sequence)	<i>DIRS</i>	
	<i>Ngaro</i>	
	<i>VIPER</i>	
PLE (<i>Penelope</i> -like element)	<i>Penelope</i>	
LINE (long interspersed nuclear element)	<i>R2</i>	
	<i>RTE</i>	
	<i>Jockey</i>	
	<i>L1</i>	
	<i>I</i>	
SINE (short interspersed nuclear element)	<i>tRNA</i>	
	<i>7SL</i>	
	<i>5S</i>	
Klasse II (DNA-Transposons) – Unterklasse I		
TIR (terminal inverted repeat)	<i>Tc1-Mariner</i>	Transposition <i>via</i> „Cut-and-Paste“-Mechanismus
	<i>hAT</i>	
	<i>Mutator</i>	
	<i>Merlin</i>	
	<i>Transib</i>	
	<i>P</i>	
	<i>PiggyBac</i>	
	<i>PIF-Harbinger</i>	
	<i>CACTA</i>	
Crypton	<i>Crypton</i>	
Klasse II (DNA-Transposons) – Unterklasse II		
Helitron	<i>Helitron</i>	Transposition <i>via</i> „Cut-and-Paste“-Mechanismus
Maverick	<i>Maverick</i>	

Tabelle 1.1: Klassifizierung transponierbarer Elemente in Eukaryoten nach Wicker et al. (2007). Die beiden Klassen werden nach ihrem Transpositionsmechanismus unterschieden.

Familien werden von der Konservierung der DNA-Sequenzen definiert. Die kleinste Einteilung erfasst eine individuelle Kopie, die zu einer spezifischen Transposition und einem Insertionsereignis gehört.

Transponierbare Elemente verlassen ihre Position im Genom. Sie integrieren sich an einer anderen Stelle, an der die Sequenz wiederholt auftritt (intramolekulare Transposition) oder ersetzen eine andere Sequenz (intermolekulare Transposition). Dabei verändern sie die Genomstruktur. Wird ein Transposon in ein Gen integriert, führt das meist zu dessen Inaktivierung. Transposons wirken sich auf die Regulation der Genexpression aus, indem sie zum Beispiel in Promotorregionen integrieren.

Transponierbare Elemente verhalten sich in filamentösen Pilzen ähnlich wie Plasmide, die für horizontalen Gentransfer zwischen Pilzen unterschiedlicher Spezies verantwortlich sind (Kistler und Miao, 1992). Sie sind in der Lage, sich innerhalb einer Population von Zellkernen zu verbreiten. *Tadl* aus *Neurospora crassa* war das erste aus einem filamentösen Pilz isolierte Transposon (Kinsey und Helber, 1989). Gene, die von der Insertion eines Transposons gestört werden, erzeugen einen selektierbaren Phänotyp. In *Neurospora crassa* wird so zum Beispiel ein Gen verändert, das für eine Glutamat-dehydrogenase codiert (Kinsey, 1977).

Horizontaler Gentransfer ist für einige transponierbare Elemente wahrscheinlicher als für andere (Silva et al., 2004). Besonders häufig ist horizontaler Gentransfer für Klasse II-Elemente vom Typ *mariner* und *P*-Elemente dokumentiert. Die Fähigkeit, sogar zwischen nicht verwandten Spezies ausgetauscht zu werden, macht transponierbare Elemente zu der beweglichsten Art von eigennütziger („selfish“) DNA. Horizontaler Gentransfer wurde bereits für verschiedene Familien transponierbarer Elemente, hauptsächlich in Metazoa (Robertson, 1995; Zhang et al., 2014), nachgewiesen. Diese Ereignisse treten oft in Wirt-Parasit-Systemen auf (Gilbert et al., 2010; Yoshiyama et al., 2001). Es ist daher möglich, dass es horizontalen Gentransfer von transponierbaren Elementen auch in dem Wirt-Parasit-System *Absidia/Parasitella* gibt. Ein Ziel dieser Arbeit war es, die Wahrscheinlichkeit von horizontalem Gentransfer innerhalb dieses genetischen Systems mit Hilfe der sequenzierten Genome von *Absidia glauca* und *Parasitella parasitica* zu ermitteln.

1.2.2 Mobile Elemente in mitochondrialer DNA

Mitochondriale DNAs von Pilzen beinhalten Homing-Endonukleasen in mobilen genetischen Elementen, meist in selbstspaltenden Introns der Gruppe I (Aguileta et al., 2014). Homing-Endonukleasen sind sehr spezifische Enzyme, die DNA gezielt spalten. Sie initiieren den Transfer des Intron-Elements, und damit ihrer eigenen Sequenz, indem sie gezielt einen Doppelstrangbruch in verwandten Allelen erzeugen, die diese intervenierende Sequenz nicht enthalten (Stoddard, 2006). Das führt zu Genkonversionen und Vererbung des mobilen Elements.

Abbildung 1.9 zeigt den generellen Mechanismus dieser hochspezialisierten transponierbaren Elemente, bei dem die Funktion des Wirtsgens nicht gestört wird.

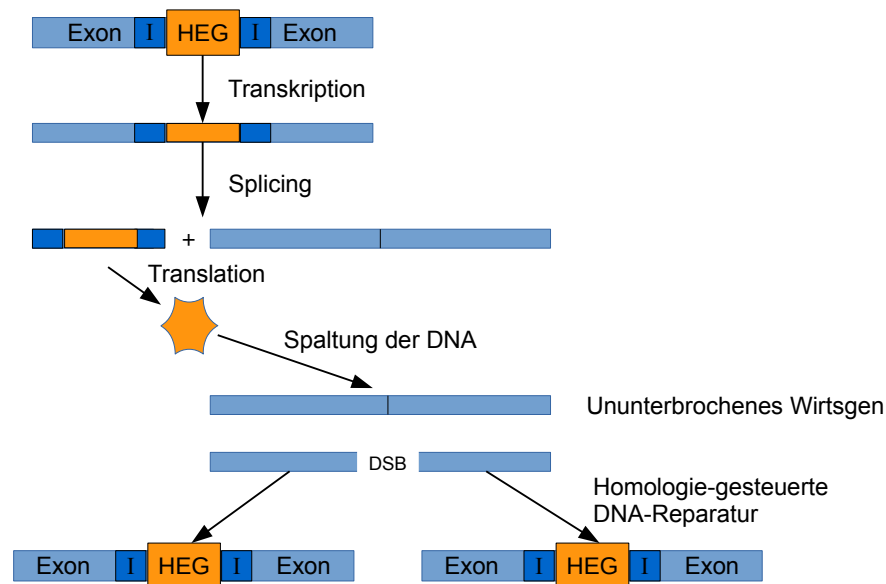


Abbildung 1.9: Mechanismus der Homing-Endonukleasen in der mtDNA von Zygomyceten. Die Endonukleaseaktivität der im Intron codierten Homing-Endonuklease erzeugt einen Doppelstrangbruch im Wirtsgen. Anhand der homologen Vorlage wird der fehlende Strang synthetisiert. Rekombination und DNA-Replikation vervollständigen den Homing-Prozess. I: Intron (Gruppe I), HEG: Homing-Endonuklease-Gen, DSB: Doppelstrangbruch (Abbildung nach Stoddard (2014))

Man unterscheidet fünf Familien von Homing-Endonukleasen nach den konservierten Kernmotiven innerhalb der aktiven Zentren der Nukleasen, den katalytischen Mechanismen, sowie der biologischen und genomischen Verteilung. Die größte bekannte Familie ist die der LAGLIDADG-Endonukleasen. Sie finden sich in Archaeen, in den Chloroplasten von Pflanzen und Algen, sowie in den Mitochondrien von Pilzen und Protozoen (Dalgaard et al., 1997). LAGLIDADG-Endonukleasen enthalten entweder ein oder zwei Kopien des konservierten LAGLIDADG-Motivs. Enzyme mit einer Kopie agieren als Homodimere und erkennen Konsensusmotive an den Zielorten in der DNA, die auf Palindrome oder näherungsweise palindromische Symmetrien beschränkt sind. Enzyme mit zwei Kopien agieren als Monomere und besitzen ein Paar strukturell ähnlicher Nukleasedomänen. Sie sind nicht auf symmetrische DNA-Ziele beschränkt. Ungebundene LAGLIDADG-Endonukleasen erkennen circa 20 bp lange DNA-Stellen (Dalgaard et al., 1994), an denen sie beide Stränge der DNA zerschneiden.

Die Familie der GIY-YIG-Endonukleasen wird von einem konservierten GIY-(X₁₀₋₁₁)-YIG-Motiv charakterisiert (Kowalski et al., 1999). Diese Endonukleasen finden sich unter anderem in den Introns mitochondrialer DNAs in Pilzen (Paquin et al., 1994). Neben dem Motiv, dem diese Familie ihren Namen verdankt, sind weitere konservierte Motive für die Faltung und Aktivität der N-terminalen katalytischen Domänen verantwortlich. Der Kern

der Nukleasen ist an eine C-terminale DNA-bindende Region gebunden. Der Aufbau und die Organisation der Domänen dieses Abschnitts der Enzyme variieren stark innerhalb der GIY-YIG-Endonukleasen (Kowalski et al., 1999).

In Pflanzen ist horizontaler Gentransfer zwischen mitochondrialen DNAs gut dokumentiert (Bergthorsson et al., 2003; Renner und Bellot, 2012; Xi et al., 2013). Der intensivste Fall von horizontalem Gentransfer findet sich in Angiospermen und betrifft ein Gruppe I Intron innerhalb der Cytochrome-c-Oxidase-Untereinheit 1 (*cox1*). Dieses Intron wurde mehrere Male während der Entwicklung der Angiospermen *via* horizontalem Gentransfer aufgenommen. Mindestens der erste Transfer ging dabei von einem Pilz aus (Vaughn et al., 1995).

1.2.3 Endosymbiotischer und horizontaler Gentransfer

Die Anzahl der bereits bekannten HGT-Ereignisse ist wesentlich höher für Ereignisse zwischen Prokaryoten, als für Ereignisse zwischen Prokaryoten und Eukaryoten, wie zum Beispiel Pilzen, Nematoden oder Protozoen (Szöllősi et al., 2015; Danchin et al., 2016; Enameh et al., 2016). Dennoch wird horizontaler Gentransfer auch für Eukaryoten als treibende Kraft in der adaptiven Evolution angesehen (Schönknecht et al., 2014). Die wesentlichsten Schritte in der Evolution von Eukaryoten, die Entwicklung von Mitochondrien und Plastiden, entstanden aus Transferereignissen über Endosymbiosen (endosymbiotischer Gentransfer, EGT) (Archibald, 2015). Die Endosymbionten, die Mitochondrien und Plastiden eukaryotischer Zellen zugrunde liegen, sind eine große Quelle an genetischem Material im Zellkern der Eukaryoten. 18 % des Genoms von *Arabidopsis* stammen von Cyanobakterien, den Vorfahren der Chloroplasten. Die Mehrheit der Gene im Zellkern stammen von α -Proteobakterien, den Vorfahren der Mitochondrien (Martin et al., 2002). Das ist auch der Grund, warum die Suche nach horizontalem Gentransfer oft den Transfer von mitochondrialen oder chloroplastidalen Endosymbionten in das Wirtsgenom identifiziert. Das macht es besonders schwierig, HGT-Ereignisse zu finden, die nicht auf diese Art von Endosymbiose zurückzuführen sind. Zum Beispiel solche, bei denen genetisches Material aus Nahrungsbakterien aufgenommen wurde. Abbildung 1.10 veranschaulicht den Unterschied zwischen endosymbiotischem (1.10 a, b) und horizontalem Gentransfer (1.10 c).

1.2.4 Etablierung endosymbiotischer Interaktionen

Nur wenige genetische Systeme, in denen horizontaler Gentransfer stattfindet, können unter Laborbedingungen untersucht werden. Auch heute noch spielt der Transfer von Genen eine wichtige Rolle, besonders bei endosymbiotischen Interaktionen in Protozoen, die eine enge Beziehung zu Bakterien haben. Oft sind gut beschriebene endosymbiotische Beziehungen allerdings schon vor Millionen Jahren entstanden.

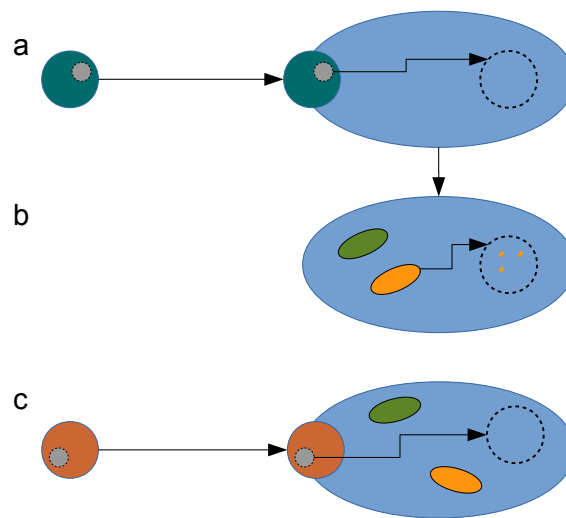


Abbildung 1.10: Veranschaulichung des Unterschieds zwischen endosymbiotischem (EGT) und horizontalem Gentransfer (HGT). Bei EGT (b) stammt das transferierte genetische Material aus einem über Endosymbiose entstandenen Organell (a), wie einem Mitochondrium (orange) oder Plastid (grün). Bei HGT ist ein eigenständiger Organismus, wie ein aufgenommenes Futterbakterium (braun), der Ursprung des Gentransfers (c). (Abbildung nach Andersson (2005))

Ein gutes Beispiel hierfür ist die Endosymbiose zwischen *Euplotes aediculatus* und *Polynucleobacter necessarius*, einem β -Proteobakterium, das essentiell für seinen Wirt ist und umgekehrt (Heckmann, 1975; Görtz, 2010). Als Konsequenz dieser intrazellulären Lebensweise haben sich die physiologischen und biosynthetischen Fähigkeiten von *Polynucleobacter necessarius* zurückentwickelt, was sich in einer drastisch reduzierten Genomgröße (1,6 Mbp) im Vergleich zu seinem nächsten frei lebenden Verwandten (2,2 Mbp) widerspiegelt (Meincke et al., 2012). Experimente, wie die Etablierung von mehr oder weniger stabilen Verbindungen zwischen einem Protozoon (*Amoeba proteus* (D-Stamm)) und einem Bakterium (Jeon, 2006), ermöglichen es, erst vor relativ kurzer Zeit entstandene Endosymbiosen in Laborversuchen zu erforschen.

Horizontaler Gentransfer von Bakterien zu Eukaryoten tritt zum Beispiel auf, wenn während des Verdauungsprozesses einzelne, als Nahrung aufgenommene Bakterien den Nahrungsvakuolen entkommen und Teile ihrer DNA zum Nukleus des Protozoons transferieren. In Experimenten mit dem Ciliaten *Tetrahymena pyriformis* und dem Bakterium *Escherichia coli* wurde genau dieses Verhalten innerhalb der Einzeller beobachtet (Siegmond et al., 2013; Dürichen et al., 2016; Wöstemeyer et al., 2016). Physikochemische Modifikationen an der Oberfläche der Nahrungsbakterien und der Einsatz synthetischer Nahrungspartikel mit chemisch definierter Oberfläche ermöglichen es diesem System, Einblick in die frühen Stadien der Entstehung von Endosymbiosen zu gewinnen. Abbildung 1.11 zeigt ein aus einer Nahrungsvakuole entkommenes Futterbakterium (Siegmond et al., 2013).

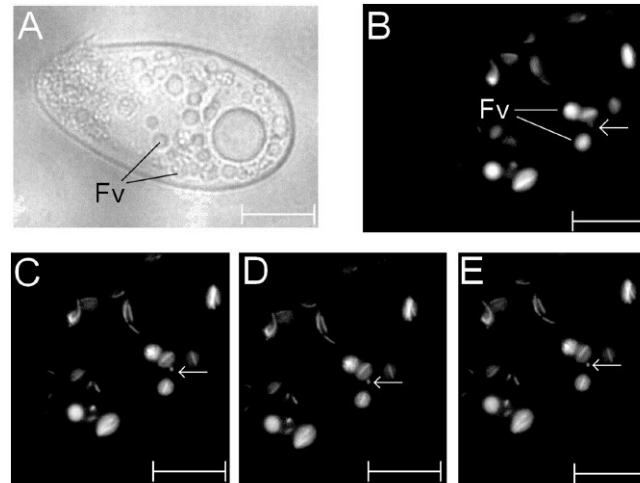


Abbildung 1.11: Mikroskopische Aufnahme 5 h nach der Fütterung mit rot fluoreszierenden *Escherichia coli* Transformanten. (A) Photographie in Differentialinterferenzkontrastierung. (B)–(E) Fluoreszenzaufnahmen. Die Pfeile markieren ein aus einer Nahrungsvakuole (Fv) von *Tetrahymena pyriformis* entkommenes Futterbakterium. Die weißen Balken entsprechen 20 µm. (Abbildung entnommen aus Siegmund et al. (2013))

Die Umgebungsbedingungen in Nahrungsvakuolen von Protozoen steigern die Konjugationshäufigkeit zwischen den aufgenommenen Bakterien (Matsuo et al., 2010; Schlimme et al., 1997). In Bakterien ist Konjugation die effektivste Methode, um Gene zu überführen. Dabei beschränkt sich der Transfer nicht auf eine enge genetische Verwandtschaft. Heinemann und Sprague (1989) haben gezeigt, dass *Escherichia coli* DNA über eine mit einem Pilus zustande kommende Verbindung zu dem Ascomyceten *Saccharomyces cerevisiae* transferiert. Dieser Vorgang ist theoretisch so ähnlich auch in Protozoen möglich.

Da es inzwischen möglich ist, Cokulturen von Protozoen und Nahrungsbakterien im Labor zu untersuchen, ergibt sich der Bedarf für neue spezifischere Detektionsmethoden zum Auffinden von HGT-Ereignissen. Wie bei der Entwicklung eines Labormodells für *Tetrahymena pyriformis* (Wöstemeyer et al., 2016), ist es nicht immer möglich, vollständige Genome aller Wirtsorganismen zu sequenzieren, die vielleicht einige interessante neue Eigenschaften zeigen, nachdem sie in Langzeitstudien zusammen mit verschiedenen Bakterien gehalten wurden. Einzelne Sequenzen reichen für die meisten etablierten HGT-Detektionsmethoden nicht aus.

Der in dieser Arbeit entwickelte Algorithmus (siehe Abschnitt 3.6) ist in der Lage, einzelne, rezente Gentransferereignisse in kleinen Datensätzen zu finden, auf die andere Algorithmen nicht anwendbar sind. Die Suche in Protozoen ist besonders interessant, da diese zum Teil von medizinischer Bedeutung sind. Horizontaler Gentransfer in *Cryptosporidium parvum* (Striepen et al., 2004), zum Beispiel, ist eine der Ursachen für die Schwierigkeiten bei der Behandlung von Cryptosporidiose.

1.2.5 Detektion von horizontalem Gentransfer

Es gibt verschiedene Techniken, um horizontalen Gentransfer nachzuweisen. Generell lassen sich diese Methoden in zwei Gruppen aufteilen (Ravenhall et al., 2015). Zum einen gibt es parametrische Ansätze, bei denen das Genom eines Organismus nach untypischen Sequenzabschnitten durchsucht wird. Eine Abweichung im GC-Gehalt liefert so einen Hinweis auf horizontalen Gentransfer. Zum anderen betrachtet man die evolutionäre Geschichte des Wirtes im Bezug auf ein bestimmtes Gen. Man verwendet dafür phylogenetische Methoden. Die meisten dieser Verfahren benötigen große Datenmengen, im schlimmsten Fall ganze Genome, Wissen über die verwandtschaftlichen Beziehungen der Organismen und daher eine große Anzahl an Taxa für einen Speziesbaum, Wissen über die Nachbarschaft des betreffenden Gens, verschiedene komplexe Algorithmen mit einer Vielzahl an Parametern, die das Ergebnis beeinflussen und viel Zeit für zusätzliche manuelle Auswertung und Editierung der erhaltenen Ergebnisse. Am Ende hat man oft nur eine Liste mit Kandidatengenen, ohne zusätzliche Informationen über deren Funktion oder den Zeitpunkt, an dem das Gen aufgenommen wurde. Oft bleibt auch der Ursprung des Gens unbekannt.

Ein Ziel dieser Arbeit war es, den Aufwand für das Auffinden von rezenten HGT-Ereignissen zu senken, indem einfache Algorithmen und kleine Datensätze für einzelne Gene verwendet werden. Ausgehend von der Annahme, dass rezente Ereignisse stärkere Signale erzeugen, die leichter zu erkennen sind, sollte endosymbiotischer Gentransfer ausgeschlossen und der Fokus auf den rezenten Gentransfer von Bakterien zu Protozoen gelegt werden.

1.3 Vorhersage von Proteinstrukturen

Im Dezember 2016 enthielt die UniProt-Datenbank (Consortium, 2015) über 71 Millionen Proteinsequenzen (www.uniprot.org/statistics/). Die Proteindatenbank PDB (Rose et al., 2015) dagegen nur über 116 000 (www.rcsb.org/pdb/statistics/holdings.do) experimentell ermittelte dreidimensionale Proteinstrukturen. Diese immer größer werdende Diskrepanz zwischen der Menge an Sequenzdaten und verfügbaren Proteinstrukturen stellt eine große Herausforderung für die Suche nach Zusammenhängen zwischen Struktur und Funktion von Proteinen dar. Fortschritte im Bereich der Vorhersage von Proteinstrukturen helfen dabei, diese Lücke zu verkleinern (Schwede, 2013). Im Durchschnitt lassen sich inzwischen 50 – 70 % der Proteinstrukturen eines typischen Genoms mit Hilfe verschiedener Ansätze vorhersagen (Mukherjee et al., 2010). Die beiden wichtigsten Annahmen für diese Methoden sind, dass die Struktur von Proteinen in deren Evolution stärker konserviert ist, als deren Sequenz und dass es Beweise für die endliche und relativ kleine Zahl (1 000 – 10 000) an einzigartigen Proteinfaltungen in der Natur gibt (Koonin et al., 2002). Diese Annahmen ermöglichen es, das Problem der Strukturvorhersage nicht

als exakte Simulation des äußerst komplexen Faltungsvorgangs zu sehen. Strukturvorhersage wird vielmehr als die Suche nach Übereinstimmungen zwischen einer gegebenen Sequenz und bereits bekannten Proteinstrukturen formuliert. Das reduziert die Komplexität der Algorithmen zur Strukturvorhersage erheblich.

1.3.1 Das Grundproblem der Proteinfaltung

Linderstrøm-Lang (1952) führte die Bezeichnungen Primärstruktur, Sekundärstruktur und Tertiärstruktur ein, um den Aufbau von Proteinen zu beschreiben. Die Primärstruktur bezeichnet dabei die eindimensionale Abfolge von Aminosäuren in einer Proteinsequenz.

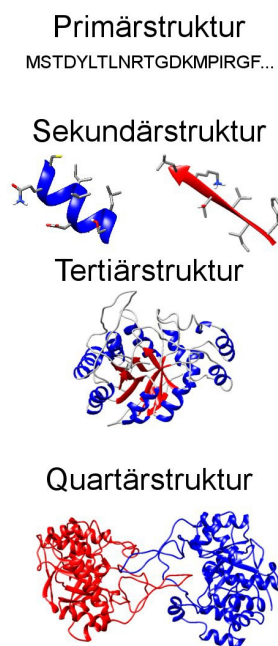


Abbildung 1.12: Aufbau von Proteinen auf verschiedenen Ebenen.

Ausgehend von der eindimensionalen Aminosäuresequenz entwickelt sich die Faltung über die Entstehung einzelner Strukturelemente hin zur Anordnung dieser im Raum. Einzelne Einheiten setzen sich zu größeren Komplexen zusammen und bilden neue Funktionen aus. Gezeigt ist die Inaktivierung einer 4-Dihydromethyltrisporat-Dehydrogenase nach Homodimerbildung. Die flexible Schleifenregion des einen Proteins verschließt das aktive Zentrum des Partnerproteins.

Sekundärstrukturelemente werden zum Beispiel im Dictionary of Protein Secondary Structure (DSSP) (Kabsch und Sander, 1983), einem Verzeichnis für Sekundärstrukturelemente in Proteinen, aufgelistet. DSSP betrachtet dabei die Torsionswinkel im Proteinrückgrat von Kristallstrukturen, Wasserstoffbrückenmuster, Lösemittelzugänglichkeit und andere Parameter, die für verschiedene Sekundärstrukturen charakteristisch sind. Man unterscheidet hier acht Arten von Sekundärstrukturen: 3_{10} -Helices mit einer minimalen Länge von drei Aminosäuren; α -Helices mit einer minimalen Länge von vier Aminosäuren; π -Helices mit einer minimalen Länge von fünf Aminosäuren; Strukturen, die mit Wasserstoffbrückenbindungen in regelmäßigen Abständen entstehen; β -Faltblätter, die aus parallel oder anti-parallel angeordneten Strängen mit einer minimalen Länge von zwei

Aminosäuren bestehen; β -Faltblätter, die aus nur einem Paar Aminosäuren bestehen und über Wasserstoffbrücken verbunden sind; β -Schleifen, die nicht auf Wasserstoffbrückenbindung zurückgehen und Coils als Bereiche, die keine definierten Sekundärstrukturen aufweisen. Sekundärstrukturelemente werden von einer Reihe von Wechselwirkungen stabilisiert. Die gefalteten Strukturen sind stabiler als die entfalteten. Die wichtigsten stabilisierenden Kräfte sind hydrophobe und ionische Wechselwirkungen.

Die Tertiärstruktur bezeichnet die räumliche Anordnung aller Atome eines Proteins. Sie beinhaltet auch Disulfidbrücken, die Positionen der Aminosäureseitenketten und alle lokalen oder globalen Wechselwirkungen. Unter dem Begriff Quartärstruktur versteht man die Zusammenlagerung mehrerer Proteine als Untereinheiten eines größeren Komplexes. Die Funktion eines Proteins ist oft an ihre Quartärstruktur (Abbildung 1.12) gebunden.

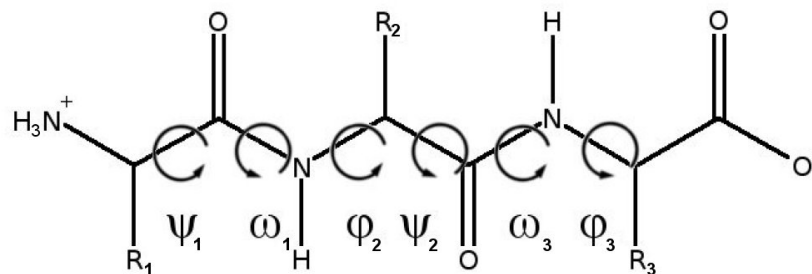


Abbildung 1.13: Torsionswinkel in Aminosäureketten.

Der Winkel φ bezieht sich auf die Bindung zwischen $N-C_{\alpha}$, der Winkel ψ auf die Bindung zwischen $C_{\alpha}-C$ und der Winkel ω auf die Peptidbindung.

Die unterschiedliche Geometrie der Sekundärstrukturelemente entsteht über die Rotation um die Einfachbindungen des C_{α} -Rückgrats der Proteine. Es gibt drei Torsionswinkel pro Aminosäure (Abbildung 1.13): φ entlang der Bindung zwischen $N-C_{\alpha}$, ψ entlang der Bindung zwischen $C_{\alpha}-C$ und ω entlang der Peptidbindung. Das Rückgrat der Proteine ist meist mehr oder minder starr. Die Seitenketten an der Oberfläche hingegen sind sehr flexibel.

Im Ramachandran-Plot (Morris et al., 1992) stellt man jeweils Paare (φ , ψ) von Torsionswinkeln einer Aminosäure dar. Bestimmte Kombinationen an Torsionswinkeln werden energetisch bevorzugt, andere sind hingegen sterisch ausgeschlossen. Sekundärstrukturelemente besitzen gut definierte Winkelkombinationen, die man im Ramachandran-Plot (Abbildung 1.14) bestimmten Regionen zuordnen kann.

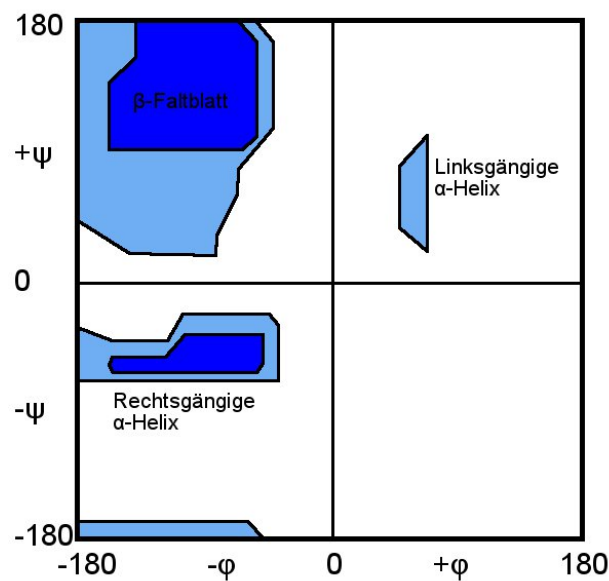


Abbildung 1.14: Im Ramachandran-Plot werden bestimmten Regionen unterschiedliche Sekundärstrukturelemente mit spezifischen Winkelkombinationen (ϕ , ψ) zugeordnet. (Abbildung nach Hollingsworth und Karplus (2010))

Seit über 50 Jahren beschäftigen sich Wissenschaftler mit der Modellierung von Proteinen (Dill und MacCallum, 2012). Dabei wirft das Problem der Proteinfaltung die folgenden drei Fragen auf:

1. Wie wird die dreidimensionale Struktur eines Proteins von den physiochemikalischen Eigenschaften der eindimensionalen Aminosäuresequenz festgelegt?
2. Wie kann ein Protein sich so schnell falten, obwohl eine Polypeptidkette eine sehr große Anzahl an möglichen Konformationen einnehmen kann?
3. Ist es möglich, mit Hilfe eines Computeralgorithmus, die Struktur eines Proteins ausgehend von einer Aminosäuresequenz ohne weitere Informationen vorherzusagen?

Das Grundproblem der Proteinfaltung besteht darin, für eine gegebene Aminosäuresequenz die passende Proteinstruktur zu finden. Dabei ist die Wahl einer geeigneten Methode abhängig von der Verfügbarkeit homologer Proteinstrukturen und experimenteller Daten, sowie der gewünschten Auflösung des Modells. Die Vorhersage berücksichtigt in erster Linie nur die Konformationen des C_α -Rückgrates. Die Seitenketten werden später modelliert.

Anfinsen führte wichtige Experimente zur Faltung durch (Anfinsen, 1973). Er zeigte, dass Mercaptoethanol die Schwefelbrücken reduktiv löst und dass Harnstoff bei niedrigem pH die Wechselwirkungen innerhalb des Proteins schwächt. Daraus resultierte eine Entfaltung des Proteins. Die Entfernung der beiden Reagenzien nach Dialyse führte

meist zum nativen Zustand zurück. So wurde der Zusammenhang zwischen Aminosäuresequenz und Proteinstruktur deutlich.

Die Biophysik der Proteinfaltung lässt sich nicht auf die eindeutige Beziehung zwischen Sequenz und Tertiärstruktur reduzieren. Gleiche Teilsequenzen finden sich in unterschiedlichen Strukturen. Unterschiedliche Sequenzen bilden allerdings auch die gleiche Tertiärstruktur aus. Das ist zum Beispiel bei Punktmutationen der Fall. Schon kleine Änderungen in der Sequenz genügen, um die Ausbildung komplett unterschiedlicher Tertiärstrukturen zu bewirken. Das Faltungsproblem ist daher nichtlokal. Anfinsens Experiment führte zur Formulierung der *Thermodynamischen Hypothese*: Faltung strebt den thermodynamisch günstigsten Zustand an. Übergangszustände sind energetisch weniger günstig, müssen allerdings während der Faltung durchlaufen werden. Der entfaltete Zustand lässt viele mögliche Konformationen zu. Die Seitenketten der Aminosäuren nehmen praktisch alle möglichen Zustände (Freiheitsgrade) an. Das erhöht die Entropie, die man daher auch Konformationsentropie nennt. Im gefalteten Zustand gibt es eine einzige Konformation des Proteinerückgrats und viele der Seitenkettenwinkel sind fest vorgegeben. Das führt zu niedriger Entropie. Faltung entspricht aus dieser Sicht dem Verlust an Freiheitsgraden. Die Kinetik der Proteinfaltung versucht, zu beschreiben, auf welchen Zeitskalen die Faltung abläuft und welche Zwischenzustände angenommen werden. Die Frage ist, welche Schritte die Geschwindigkeit bestimmen und ob sich Sekundär- und Tertiärstruktur gleichzeitig ausbilden.

Das Levinthal-Paradoxon (Levinthal, 1968) beschreibt das Problem dieser kombinatorischen Explosion bei der Faltung eines Proteins. Ausgehend von der Annahme, dass die komplette Aminosäurekette ihre physiologische dreidimensionale Form erst sucht, nachdem sie vollständig synthetisiert wurde, muss eine extrem hohe Zahl möglicher Konformationen in einer zufälligen Suche ausprobiert werden. Die Auflösung des Levinthal-Paradoxons ergibt sich aus dem Umstand, dass die Suche nicht zufällig abläuft. Tatsächlich nimmt jede Aminosäure nach Aufnahme in die Sequenz die energetisch günstigste Raumrichtung ein, wofür die Zeit bis zum Anfügen der nächsten Aminosäure ausreichend ist. Benachbarte Sequenzabschnitte falten sich zu Sekundärstrukturelementen des nativen Proteins. Diese werden *foldon* genannt (Hu et al., 2013). Jedes gefaltete Segment bleibt während dem schrittweisen Hinzufügen weiterer Elemente erhalten. Auf diese Weise wird die native Proteinstruktur Stück für Stück und einem einzelnen dominanten Reaktionsweg folgend aufgebaut.

Der Faltungsprozess entspricht einer Trajektorie im Konformationsraum, die zum globalen Minimum hin verläuft. Um das Minimum zu erreichen, sind unterschiedliche Startkonformationen möglich. Abbildung 1.15 zeigt die trichterförmige Energielandschaft einer Proteinfaltung. Nicht gefaltete Proteinstrukturen führen zu einem höheren Energieniveau, bereits gefaltete Proteinstrukturen ergeben die wenigen niedrigeren Energieniveaus (Dill und MacCallum, 2012).

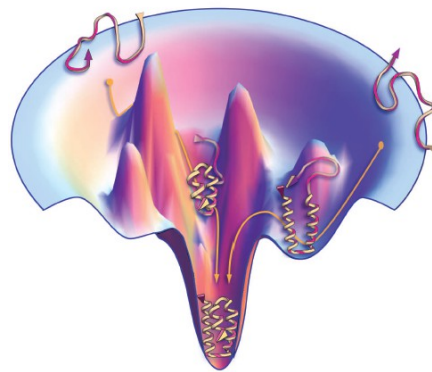


Abbildung 1.15: Trichterförmige Energielandschaft von Proteinen.

Nicht gefaltete Proteinstrukturen führen zu einer Vielzahl an höheren Energieniveaus, bereits gefaltete Proteinstrukturen ergeben die wenigen niedrigeren Energieniveaus.

(Abbildung entnommen aus Dill und MacCallum (2012))

Aus dem Versuch, mit Proteinstrukturvorhersage den zeit- und kostenaufwändigen Prozess der experimentellen Strukturaufklärung zu umgehen, sind inzwischen weitere anspruchsvolle Anwendungsbereiche hervorgegangen. Die Entwicklung künstlicher Proteinstrukturen (Kirubakaran et al., 2016; Nastro et al., 2016) ist ebenso möglich, wie die Suche nach neuen Substanzen in der Wirkstoffforschung mit Liganden-Screening (Aldib et al., 2016; Surekha et al., 2017).

1.3.2 Empirische Strukturvorhersage

Eine Möglichkeit, die Struktur eines Proteins empirisch vorherzusagen, ist der Chou-Fasman-Algorithmus (Chou und Fasman, 1978). Die Idee hinter diesem Algorithmus ist die Ermittlung statistischer Unterschiede in der Neigung der Aminosäuren zur Ausbildung bestimmter Sekundärstrukturen. Aus Strukturdatenbanken wie PDB (Rose et al., 2015) lässt sich ermitteln, wie oft welche Aminosäure in welcher Sekundärstruktur auftritt. Die 20 Aminosäuren werden danach in unterschiedliche Klassen eingeteilt: sehr häufig in Helices vertreten (Glutaminsäure, Methionin, Alanin, Leucin), durchschnittlich in Helices vertreten (Lysin, Phenylalanin, Glutamin, Tryptophan, Isoleucin, Valin), Indifferente (Asparaginsäure, Histidin, Arginin, Threonin, Serin, Cystein), schwache Helixbrecher (Tyrosin, Asparagin) und starke Helixbrecher (Prolin, Glycin). Eine analoge Einteilung gibt es für Faltblätter. Leider ist die Vorhersagegenauigkeit dieses eher einfachen Algorithmus gering (50 – 60 %). Nagano (1973) entwickelte eine statistische Methode zum Auffinden von Helices, β -Faltblättern und Schleifen in Proteinen. Auch mit dieser Methode lag die Vorhersagegenauigkeit bei 52 – 64 %.

Zwei Proteine aus der Arbeit von Nagano (1973) verdeutlichen ein weiteres Problem bei der Vorhersage von Proteinfaltung. Es ist möglich, dass dieselbe Peptidsequenz unterschiedliche Sekundärstrukturen ausbildet. Ein Beispiel hierfür ist die Aminosäuresequenz Val-Asn-Thr-Phe-Val, die sowohl in Erythrocyruorin (PDB: 1ECN) (80-84, Helix) als auch

in der Ribonuklease A (PDB: 9RSA) (43-47, Faltblatt) vorkommt. Um diese Nichtlokalität zu berücksichtigen, werden bei anderen Methoden benachbarte Reste in die Vorhersage miteinbezogen. Diese Vorgehensweise verbessert die Vorhersage für Helices deutlich. Die Vorhersage von Faltblättern bleibt immer noch schwierig, da für die Stabilisierung von Faltblättern Wechselwirkungen zwischen entfernten Sequenzbereichen notwendig sind. Viele Methoden zur Vorhersage von Proteinstrukturen basieren auf künstlichen neuronalen Netzen, Linearer Diskriminanzanalyse, Nächster-Nachbar-Klassifizierern, Support-Vektor-Maschinen oder Hidden-Markov-Modellen.

1.3.3 Homologie-basierte Strukturvorhersage

Das Vergleichen einer Sequenz mit Sequenzen, deren Proteinstrukturen bereits bekannt sind, bezeichnet man als Homologiemodellierung oder *Threading*. Dabei werden Sequenz-Struktur-Alignments erzeugt. Man unterscheidet eine Reihe von Faltungsklassen. Es gibt Proteine, die nur aus α -Helices oder nur aus β -Faltblättern bestehen. Andere Proteine sind aus einer sich wiederholenden Abfolge dieser beider Strukturelemente aufgebaut, wie zum Beispiel in TIM-Barrel-Proteinen. Hier ist das Grundmuster $\beta - \alpha - \beta$, das heißt, zwei Stränge werden über eine Helix miteinander verbunden. Stränge und Helices liegen innerhalb eines Proteins auch getrennt voneinander.

Konvergente Evolution beschreibt die Zuordnung zweier Proteine zu derselben Faltungsklasse, ohne dass diese evolutionär miteinander verwandt sind. Divergente Evolution führt zu funktionellen Gemeinsamkeiten und somit zu ähnlichen strukturellen Lösungen. Die Ähnlichkeiten der Strukturen sind auf kleine Fragmente, wie zum Beispiel ein aktives Zentrum oder ähnliche funktionelle Bereiche, beschränkt. Proteine, die evolutionär verwandt sind, weisen aufgrund divergenter Evolution häufig nur sehr geringe Sequenz-ähnlichkeiten auf. Da die Anzahl der Faltungsklassen beschränkt ist, kommt es auch vor, dass zwei Proteine durch Zufall derselben Faltungsklasse zugeordnet werden. Probleme bereiten die flexiblen Schleifenregionen zwischen den Strukturelementen, die keiner festen Struktur folgen und in denen sich Proteine einer Faltungsklasse unterscheiden. Für die Strukturvorhersage muss eine geeignete Energiefunktion gefunden werden. Bei schlechter atomarer Auflösung der Strukturen ist das ein schwieriger Prozess.

Wissensbasierte Energiefunktionen werden aus der Analyse von Strukturdatenbanken abgeleitet. Die inverse Boltzmann-Statistik beschreibt die Energie aus der Häufigkeit des Auftretens in der Datenbank. So etwas nennt man Pseudo-Energien, da keine physikalischen Modelle notwendig sind, um die Energie abzuleiten. Ein Beispiel ist das Sippl-Potential (Sippl, 1990). Der Vorteil von Paar-Interaktionspotentialen liegt in der geringen rechnerischen Komplexität. Sie ist auf das Auslesen eines Tabelleneintrags reduziert. Man benötigt kein physikalisches Modell oder Modellierung der Seitenketten. Man erhält so nur eine sehr grobe Näherung, da die einzelne Aminosäure als Einheit betrachtet wird. Viele spezifische und seltener auftretende Effekte sind mit dieser Methode nicht erfassbar.

Das Verfahren ist von guten Strukturdatenbanken abhängig, die man zur Parametrisierung benötigt.

Wenn man neue Sequenzen an bereits bekannte Faltungen anlegt, um sie zu modellieren, lassen sich neue noch unbekannte Faltungen nicht identifizieren. Das ist ein entscheidender Nachteil dieser Herangehensweise. Oft ist auch nur der Kern eines Proteins so gut konserviert, um korrekte Vorlagen-basierte Vorhersagen zu ermöglichen. Die Länge und der Startpunkt der einzelnen Strukturelemente variieren und erschweren die Modellierung. Threading ist im Allgemeinen ein NP-schweres Problem. Da Threading das inverse Faltungsproblem darstellt, ergeben sich ähnliche Schwierigkeiten. Die Anzahl möglicher Alignments ist exponentiell. Die Potentialfunktion ist nichtlokal. Die Wechselwirkung beliebig angeordneter Aminosäuren miteinander innerhalb der Sequenz macht paarweise Potentiale für Threading ungünstig. Alignments ohne Lücken resultieren in einer linearen Anzahl an Alignments. Einfachere Potentialfunktionen lösen die Nichtlokalität auf und das Alignmentproblem wird zu einem gewöhnlichem Profilalignment, das mit dynamischer Programmierung in polynomieller Zeit lösbar ist.

Solvatationspotentiale beschreiben die Lösemittelzugänglichkeit eines Proteins, das heißt, wie stark der Aminosäurerest für Wasser zugänglich ist. Das entspricht dem Anteil der Oberfläche der Aminosäure, der Kontakt mit Wasser hat. Der Wert geht von null: unzugänglich im Inneren des Proteins bis eins: komplett zugänglich auf der Proteinoberfläche. Die Solvation ist wichtig für die Proteinfaltung. Jones et al. (1992) leiten das Solvationspotential ähnlich dem Sippl-Potential (Sippl, 1993a) her.

Kontaktkapazitätspotentialen liegt die Idee zu Grunde, dass Zielstruktur und Vorlage eine identische Faltung aufweisen und damit auch eine ähnliche Biochemie, zum Beispiel Hydrophobizität. Hydrophobe Bereiche in der Vorlage sollten von hydrophoben Aminosäuren in der Zielsequenz ausgefüllt werden. Die Herleitung ist ähnlich, wie die der Kontaktpotentiale. Es werden die Aminosäurereste in der Nähe einer Position der Vorlage gezählt, unabhängig vom Typ des Aminosäurerests an dieser Position in der Vorlage. Prinzipiell entsteht so ein abgeleitetes Profil für die Aminosäurepräferenz an einer bestimmten Position aus der Vorlage. Das Profil ist dadurch nicht von der Ausgangssequenz abhängig.

1.3.4 Strukturvorhersage mit künstlichen neuronalen Netzen

Da nur etwa 65 % der Information lokaler Natur sind, lassen sich die Methoden der ersten und zweiten Generation, wie zum Beispiel der Chou-Fasman-Algorithmus (Chou und Fasman, 1978), kaum verbessern.

50 – 80 % der Aminosäuren eines Proteins sind austauschbar, ohne die Sekundärstruktur des Proteins entscheidend zu verändern (Rost, 1999; Sander und Schneider, 1991; Shakhnovich und Gutin, 1991). Im Laufe der Evolution fanden viele dieser neutralen Mutationen statt. Evolutionär homologe Proteinsequenzen enthalten diese Information.

Treten an einer Position in homologen Sequenzen häufig Helixbrecher auf, ist es unwahrscheinlich, dass dort eine Helix liegt. Diese Art von Information lässt sich auf einfache Art in Form von Profilen einbringen.

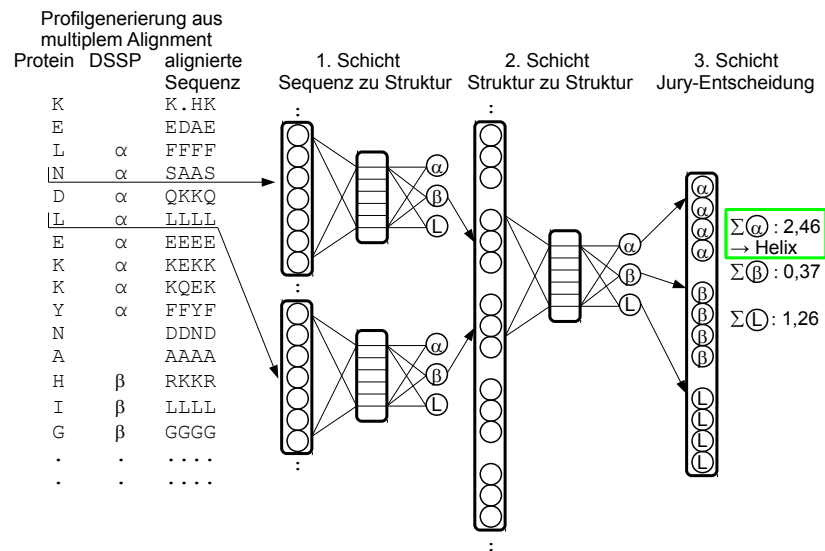


Abbildung 1.16: Anwendung künstlicher neuronaler Netze zur Vorhersage von Proteinstrukturen. (Abbildung nach Rost und Sander (1993))

PSIPRED (Jones, 1999) ist ein dreistufiger Algorithmus zur Vorhersage von Proteinstrukturen unter Verwendung eines zweistufigen künstlichen neuronalen Netzwerks (ANN, Artificial Neural Network). Dieses interpretiert ein mit PSI-BLAST (Altschul et al., 1997) erzeugtes Profil-Alignment. Die Proteinsequenz wird in einer großen, nicht-redundanten Proteinsequenz-Datenbank gesucht und die Ergebnisse als PSI-BLAST-Profil (Scoring-Matrix) gespeichert. Diese Matrix ist die Eingabe für die erste Schicht des neuronalen Netzwerks. In der zweiten Schicht werden die Strukturen untereinander verglichen. In Schicht drei erfolgt die Zuordnung zu einer der drei Strukturelemente Helix, Strang oder Schleife. Diese Zuordnung erfolgt aufgrund eines Mehrheitsentscheids. Abbildung 1.16 zeigt den generellen Aufbau eines neuronalen Netzes zur Vorhersage von Proteinstrukturen nach Rost und Sander (1993).

1.3.5 De novo Strukturvorhersage

Bei der *de novo* oder auch *ab initio* Modellierung werden Vorhersagen ausgehend von physikalischen Modellen nur anhand der Aminosäuresequenz eines Proteins getroffen. Dabei werden keine homologen Strukturen verwendet. Der Vorteil dieses Ansatzes ist, dass so auch die Vorhersage neuer noch unbekannter Faltungen möglich ist. Der Nachteil besteht in hoher Komplexität und Rechenaufwand, die die Anwendbarkeit auf kleine Proteine (weniger als 100 Aminosäuren) beschränken.

Das gefaltete Protein nimmt innerhalb der Energielandschaft ein Minimum ein. Man

verwendet Algorithmen zur Energieminimierung und Durchmusterung des Konformationsraums. Das Problem dabei ist, dass Proteine oft mehrere Faltungspfade kennen. Sie haben unterschiedliche Möglichkeiten, zu ihrer endgültigen Faltung zu gelangen. Die Faltungsgeschwindigkeit verschiedener Domänen ist unterschiedlich. Die Proteinstruktur selbst ist nicht lokal, sondern von der Umgebung abhängig. Identische Aminosäureabfolgen finden sich in unterschiedlichen Sekundärstrukturen.

Eine Möglichkeit, Proteinfaltung zu studieren, ist das Gittermodell. Dieses Modell ist sehr stark vereinfacht. Ein Protein entspricht einer Kette von Kugeln, die die Aminosäuren repräsentieren. Die Kugeln liegen dabei immer auf den Gitterpositionen. Zwischen benachbarten Kugeln kommt es zur Ausbildung von Wechselwirkungen. Eine Variante ist das HP-Modell (Lau und Dill, 1989), bei dem das Protein als eine Sequenz hydrophober und polarer Reste dargestellt wird (Abbildung 1.17). Wechselwirkungen kommen nur zwischen benachbarten, nicht direkt verbundenen Resten zustande. Trotz grober Vereinfachung enthält das HP-Modell viele Eigenschaften des Faltungsproblems. Zum Beispiel das Levinthal-Paradoxon und die Größenordnung der Faltungsgeschwindigkeit. Da dem Modell eine Energiefunktion zugrunde liegt, lassen sich interessante Aussagen über die Komplexität des Problems und die Struktur der Energielandschaften ableiten. Gesucht wird die Anordnung der Sequenz auf dem Gitter mit minimaler Energie, wobei alle möglichen Anordnungen nur für sehr kurze Sequenzen (um die 30 Aminosäuren) berechenbar sind. Proteinfaltung ist im HP-Modell für alle sinnvollen Gittertypen mit Energiefunktionen, die ähnlich zu van-der-Waals-Wechselwirkungen sind, NP-hart. Approximationen sind in linearer Zeit möglich.

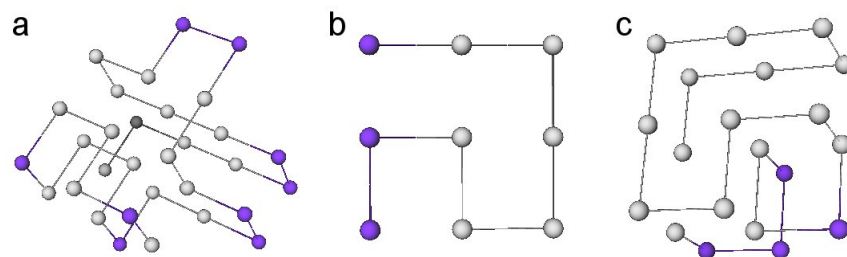


Abbildung 1.17: HP-Modelle zur Vorhersage von Proteinstrukturen.

Die Proteinsequenz wird auf eine Abfolge hydrophober (grau) und hydrophiler (polarer, lila) Aminosäurereste reduziert. Die drei Modelle stellen einzelne Abschnitte des TSP1 CEP13671 aus *Parasitella parasitica* dar, (a) interne Schleifenregion, (b) α 5-Helix und (c) C-terminales Ende des Proteins. Die Modelle wurden mit HPstruct erzeugt (Mann et al., 2009).

Die Simulation von Moleküldynamiken (MDS) erfasst die Faltung von Proteinen als Prozess, sowohl in Abhängigkeit von der Energie, als auch von der Dynamik der Aminosäuren. Die Herausforderung hierbei besteht in der Präzision der Energiefunktion, bei der das globale Minimum erhalten bleiben muss, und im Aufwand, der nötig ist, um die Faltung zu realisieren. Kleinere Proteine falten im Bereich von μ s, benötigen im Modell hingegen $10^8 - 10^{10}$ Iterationen.

Die MDS eines kleinen Peptids (36 Aminosäuren) von Duan et al. (1998) verwendete AMBER (Assisted Model Building with Energy Refinement) auf einem Computercluster (CRAY T3D) mit 256 CPUs. Die Simulationszeit betrug 200 ns (10^8 Iterationen) bei konstanter Temperatur und konstantem Druck. Dabei wurde selbst nach mehreren Monaten Rechenzeit der native Proteinzustand nicht erreicht.

Ein neuerer Ansatz beschleunigt das Ganze um das Vierfache und läuft auf einem normalen PC (Shen und Freed, 2002).

AMBER berücksichtigt fünf Parameter: Bindungslänge, Bindungswinkel, Torsionen, van-der-Waals-Wechselwirkungen und Elektrostatik. Es ist nicht nur für Proteine, sondern auch für DNA geeignet und wurde zur Konstruktion von Modellen für Röntgenkristallographie, später auch Kernspinresonanzspektroskopie entwickelt. Zu berücksichtigen ist, dass es nicht nur ein AMBER-Kraftfeld gibt, sondern ganze Familien, die üblicherweise mit dem Jahr der Publikation benannt werden (etwa AMBER89, AMBER94, AMBER99).

Es gibt verschiedene Kraftfeldklassen, die nach den von ihnen modellierten Wechselwirkungen unterschieden werden. Kraftfelder der Klasse I (AMBER, CHARMM, GRO-MOS) reproduzieren die optimale Geometrie recht gut und benutzen nur harmonische Terme, keine Kreuzterme. Kraftfelder der Klasse II (MM2, MM3, MMFF94) dienen der Vorhersage von spektroskopischen Eigenschaften und benutzen Terme dritten und höheren Grades, sowie Kreuzterme. Kraftfelder der Klasse III beinhalten zusätzlich die Integration chemischer Eigenschaften und Effekte, wie zum Beispiel Elektronegativität oder Hyperkonjugation.

Ein neuer Ansatz zur *de novo* Proteinstrukturvorhersage ist UniCon3D (Bhattacharya et al., 2016). UniCon3D synthetisiert und assembliert foldon-Untereinheiten (Hu et al., 2013) schrittweise über ein neuartiges Wahrscheinlichkeitsmodell, das lokale Konformationsabweichungen sowohl des Proteinrückgrats als auch der Seitenketten gleichzeitig berücksichtigt.

2 Material und Methoden

An den Stellen im Text, an denen keine Stammnummer angegeben wird, ist mit *Mucor circinelloides* immer der Stamm CBS 277.49 und mit *Parasitella parasitica* der Stamm CBS 412.66 NGI-315 ade^- gemeint.

2.1 Genomsequenzierung, -assemblierung und -annotierung

Die Genome des mit *Mucor* verwandten Fusionsparasiten *Parasitella parasitica* (CBS 412.66 NGI-315 ade^-) und dessen Wirts *Absidia glauca* (CBS 101.48 RVII-324 met^-) wurden mit Illumina MiSeq und Roche 454 sequenziert (Eurofins Genomics, Ebersberg, Deutschland). Bei diesem Verfahren handelt es sich um Schrotschuss-Sequenzierung (shotgun sequencing). Das Ergebnis einer Schrotschuss-Sequenzierung ist nicht eine einzelne Sequenz, die das gesamte Genom darstellt, sondern Millionen von Bruchstücken (Reads), die erst zu längeren Einheiten (Contigs, Scaffolds) zusammengesetzt werden müssen. Reads sind einzelne oder zu Paaren angeordnete Basensequenzen mit einer durchschnittlichen Länge von 150 – 300 Nukleotiden. Ein Contig ist eine Konsensussequenz von Reads, die sich in einem multiplen Sequenzalignment überlappen. Im Idealfall wird auf diese Weise das vollständige Genom zusammengesetzt. Eine Herausforderung stellen dabei repetitive Sequenzabschnitte dar, die das eindeutige Zusammensetzen der einzelnen Reads erschweren. Ein Genom ist selten eine einzelne Sequenz, außer zum Beispiel bei Bakterien. In Eukaryoten ist die genetische Information auf Chromosomen angeordnet. Scaffolds enthalten Contigs in einer mit Hilfe der paired-end Reads definierten Reihenfolge und Orientierung. Bei guter Sequenzierung und Assemblierung entsprechen sie ganzen Chromosomen. Häufig gibt es allerdings Bereiche, die nicht von einem Contig abgedeckt und als Lücken gekennzeichnet werden. Nicht alle Reads kommen mit derselben Häufigkeit innerhalb der Bibliotheken vor. Mitochondriale Sequenzen treten zum Beispiel häufiger auf, als Gene, die nur einmal im gesamten Genom vorkommen (single copy genes).

Abbildung 2.1 zeigt die Assemblierung eines Scaffolds ausgehend von single-end und paired-end Reads. Für die Assemblierung der Scaffolds (Eurofins Genomics, Ebersberg, Deutschland) wurden zwei Bibliotheken mit paired-end Reads im Abstand von 3 kb beziehungsweise 8 kb verwendet. Die Reads wurden mit Trimmomatic (v0.32) (Bolger et al., 2014) analysiert und beschnitten. Adaptersequenzen und Basen mit schlechter Qualität (Phred-Wert < 20) am Anfang oder Ende eines Reads wurden entfernt. Sequenzen, die nicht den durchschnittlichen Qualitätsschwellwert innerhalb eines 4 bp langen Fensters erfüllten, wurden ebenfalls beschnitten. Zum Schluss wurden alle Sequenzen, die kürzer als 150 bp waren, entfernt. Das Genom wurde mit Newbler (v2.9) und Velvet (v1.2.10) (Zerbino und Birney, 2008) assembliert.

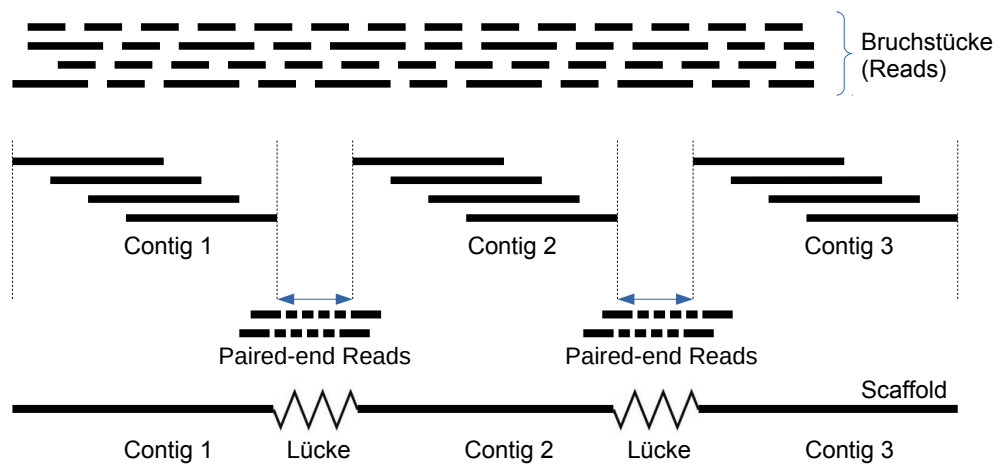


Abbildung 2.1: Schritte bei der Assemblierung eines Genoms. Die single-end Reads überlappen sich und werden zu neuen Sequenzeinheiten (Contigs) verlängert. Mit Hilfe von Reads, die zu Paaren angeordnet sind (paired-end) und deren Abstand zueinander bekannt ist, werden die einzelnen Contigs zu Scaffolds zusammengesetzt. Bereiche, die nicht von einem Contig abgedeckt werden, werden als Lücke gekennzeichnet.

Die Gene der beiden sequenzierten Genome wurden mit AUGUSTUS (v3.0.2) (Stanke und Morgenstern, 2005) vorhergesagt. AUGUSTUS basiert auf Hidden-Markov-Modellen (HMMs). Diese Modelle definieren Wahrscheinlichkeitsverteilungen für variierende Abschnitte einer genomischen Sequenz. Introns, Exons und Bereiche zwischen den Genen entsprechen den Zuständen im Modell. Jeder Zustand erzeugt DNA-Sequenzen mit einer vordefinierten Wahrscheinlichkeit. Dabei müssen bestimmte zum Teil biologisch relevante Regeln beachtet werden. Ein Exon darf zum Beispiel kein Stopcodon in seinem Leserahmen besitzen. Schnittstellen zwischen Exons und Introns müssen dem Konsensusmotiv GT–AG folgen. Alle vollständigen Gene müssen mit ATG (Startcodon) beginnen und mit einem Stopcodon enden. Alle Gene müssen enden, bevor ein neues Gen beginnen kann. Exons und Introns müssen von der Spezies abhängig definierte Mindestlängen erreichen. Weitere Beschränkungen werden zusätzlich als GFF-Datei (www.ncbi.nlm.nih.gov/Sitemap/samplerecord.html) definiert, die das Programm dazu zwingen, bestimmte Vorhersagen zu machen oder zu verwerfen. Das Programm wurde mit *Rhizopus oryzae* als ausreichend nah verwandtem Referenzorganismus gestartet. Die Gene wurden mit BLASTP (Altschul et al., 1990) in NCBI (www.ncbi.nlm.nih.gov/) und InterProScan (Jones et al., 2014) annotiert.

Um repetitive Elemente zu finden, wurden RepeatModeler und RepeatMasker (<http://repeatmasker.org>) verwendet. RepeatMasker sucht nach vorgegebenen DNA-Sequenzen in einer von Repbase (www.girinst.org/repbase/) abgeleiteten Bibliothek, die repetitive Elemente beinhaltet, und Dfam (www.dfam.org/), einer Datenbank, die Alignments und HMMs von Transposon-Elementen und anderen repetitiven DNA-Elementen sammelt. Zusätzlich wurde mit RepeatModeler eine eigene Datenbank direkt aus dem Genom erzeugt, die mögliche Spezies-spezifische repetitive Sequenzen enthält. Diese wurden von

RepeatMasker im Genom lokalisiert. RepeatModeler ist ein *de novo* Identifikator für Repeatfamilien, der zwei Programme zum Auffinden von Repeats verwendet, RECON und RepeatScout. Repeats und deren familiäre Zusammenhänge werden direkt aus den eigenen Sequenzdaten ermittelt. RepeatModeler unterstützt das automatisierte Verfahren für RECON und RepeatScout für eine gegebene genomische Datenbank und verwendet deren Ausgabe, um Konsensusmodelle putativer Repeats zu erzeugen, zu verfeinern und zu klassifizieren.

Die mitochondrialen Sequenzen von *Absidia glauca* und *Parasitella parasitica* wurden mit einer erhöhten Sequenz-Abdeckungsrate und einem niedrigen GC-Gehalt einzelner Scaffolds identifiziert. Mögliche offene Leserahmen (open reading frames, ORFs) wurden mit Artemis (Rutherford et al., 2000) definiert und annotiert. Verwendet wurde dabei der mitochondriale genetische Code für Schimmelpilze (<https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi#SG4>). Die mögliche Funktion der Gene wurde mit NCBI BLASTP (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) ermittelt. Introns und rRNAs wurden mit BLASTN und paarweisen Vergleichen zu mtDNAs anderer Pilze gefunden. tRNAscan-SE (Schattner et al., 2005) wurde dazu verwendet, um tRNAs zu identifizieren. Die mitochondriale RNase P (*rnpB*) wurde mit Mustersuche nach zwei konservierten Regionen (CRI und CRV) aus bereits bekannten *rnpB*-Genen gefunden (Seif et al., 2003). Das *rps3*-Gen (Protein der kleinen ribosomalen Untereinheit 3), das in den mtDNAs der meisten Zygomyceten fehlt, wurde mit einer BLASTP-Suche nach dem *rps3* aus *Mortierella verticillata* (UniProt: A0A086VL80) im Kerngenom dieser Pilze gefunden. Multiple Sequenzalignments von *rnpB* und *rps3* wurden mit Clustal Omega (Sievers et al., 2011) erzeugt. Das lokale Alignment der konservierten Bereiche in *rnpB*-Genen (Abbildung 3.4) wurde per Hand zusammengestellt. Die lineare Darstellung der mitochondrialen Genome (Abbildungen 3.2, 3.3) wurden mit OrganellarGenomeDRAW (Lohse et al., 2013) erstellt.

2.2 Mitochondriale DNA weiterer Zygomyceten

Die vollständig annotierten mitochondrialen DNAs (mtDNAs) von *Glomus sp.* DAOM 240422 (Glosp; KC164355; Beaudet et al., 2013), *Lichtheimia ramosa* (Licra; KJ561171; Leung et al., 2014), *Mortierella verticillata* NRRL 6337 (Morve; NC_006838; Seif et al., 2005), *Mucor circinelloides* CBS 277.49 (Mucci; KR809877), *Phycomyces blakesleeanus* NRRL 1555 (Phybl; NC_027730), *Rhizopus oryzae* DAOM 148428 (Rhior; NC_006836; Seif et al., 2005), und *Zancudomyces culisetiae* 18-3 (Zancu; NC_006837; Seif et al., 2005) stammen aus GenBank (www.ncbi.nlm.nih.gov/). Die mtDNAs von *Absidia repens* NRRL 1336 (Absre), *Backusella circina* FSU 941 (Bacci) und *Rhizopus microsporus* ATCC 11559 (Rhimi), sowie einer zweiten *Parasitella parasitica* NRRL 2501 (Parpar) stammen aus der JGI MycoCosm-Datenbank (Grigoriev et al., 2014) und wurden per

Hand annotiert. Die mitochondrialen DNAs von *Lichtheimia corymbifera* FSU 9682 (Licco) in GenBank (www.ncbi.nlm.nih.gov/) und *Umbelopsis ramanniana* AG (Umbra) in JGI MycoCosm (Grigoriev et al., 2014) wurden mit einer Suche nach mitochondrialen Genen aus *Rhizopus oryzae* (NC_006837) identifiziert und ebenfalls per Hand annotiert. Dazu wurde TBLASTN in NCBI (blast.ncbi.nlm.nih.gov/Blast.cgi) und JGI MycoCosm (genome.jgi.doe.gov/pages/blast-query.jsf?db=Umbra1) verwendet. JGI MycoCosm beinhaltet nur Scaffolds, die länger als 1 000 Nukleotide sind. Deshalb ist es nicht möglich, das Chondriom von *Umbelopsis ramanniana* zu einer einzigen Sequenz zu assemblieren. Die 12 Scaffolds wurden zu sieben zusammenhängenden Sequenzen verbunden. Auf diesen sind alle mitochondrialen Gene zu finden. Die Chondriome von *Rhizopus microsporus* und der zweiten *Parasitella parasitica* sind ebenfalls keine einzelnen Sequenzen. In der JGI MycoCosm-Datenbank bestehen die mtDNAs aus jeweils zwei Scaffolds.

2.3 Identifizierung von TSP-Proteinen

Neue Sequenzen für 4-Dihydromethyltrisorat-Dehydrogenasen (TSP1), 4-Dihydrotrisorin-Dehydrogenasen (TSP2) und Carotin-Oxygenasen (TSP3) wurden mit einer BLAST-Suche nach bereits identifizierten TSP-Proteinen ermittelt und nach Überprüfung der Übereinstimmung von Aminosäuren an entscheidenden Stellen, wie dem aktiven Zentrum oder der Bindetasche für NADP, als TSP-Proteine bestätigt. Für TSP1 und TSP2 wurden die TSP-Sequenzen aus *Mucor mucedo* verwendet (UniProt: Q01213, B9DR41), für TSP3 die TSP3-Sequenz aus *Blakeslea trispora* (UniProt: A5HXP5). Neben JGI MycoCosm (Grigoriev et al., 2014) wurden auch NCBI (www.ncbi.nlm.nih.gov/) und UniProt (Consortium, 2015) als Datenbanken verwendet.

Das Molekulargewicht der TSP-Proteine wurde mit dem Protein Molecular Weight Calculator von science gateway (www.sciencegateway.org/tools/proteinmw.htm) berechnet.

Die multiplen Sequenzalignments der TSP-Proteine wurden mit Clustal Omega (Sievers et al., 2011) erzeugt. Die Neighbor-Joining-Bäume der TSP-Proteinsequenzen (Abbildungen 3.9, 3.22 und 3.29) wurden mit Hilfe von FigTree 1.4.2 (tree.bio.ed.ac.uk/software/figtree/) aus dem von Clustal Omega im Newick-Format ausgegebenen Baum erstellt. Für den TSP1-Baum wurden Sequenzen aus zwei Pilzen, die nicht zu den Zygomyceten gehören, als Außengruppe verwendet, um den Baum zu wurzeln.

Die Sequenz-Logos für die C-terminalen Motive der TSP1-Sequenzen (Abbildung 3.10) wurden mit Web-Logo (V2.8) (Crooks et al., 2004) unter Verwendung des Farbschemas für Clustal (www.jalview.org/help/html/colourSchemes/clustal.html) erstellt.

2.4 Homologiemodellierung

Homologiemodellierung konstruiert ein vollständiges Proteinmodell auf der Grundlage einer Vorlagenstruktur. Diese wird mit Sequenz-Profil-Abgleich ermittelt.

In der Arbeit von Rost (1999) wurde gezeigt, dass Homologie bei rund 90 % der betrachteten Sequenzpaare auftrat, wenn deren Sequenzidentität bei über 30 % lag. Für Sequenzpaare mit einer Sequenzidentität von weniger als 25 % lag der Anteil homologer Sequenzen nur bei 10 %. Vorlagen für Homologiemodellierung sollten daher eine Sequenzidentität von mehr als 25 % mit der zu modellierenden Sequenz liefern.

Phyre2 (Kelley et al., 2015) ist einer der am häufigsten verwendeten Server zur Vorhersage von Proteinstrukturen. Pro Jahr wird der Server von rund 40 000 verschiedenen Anwendern benutzt. Am Tag werden 700 – 1 000 Proteinanfragen bearbeitet. In Zusammenarbeit mit anderen Gruppen wurde Phyre2 bereits zur Annotation von Genomen eingesetzt. Das System läuft auf einem Linux-Cluster mit circa 300 CPUs.

Grundsätzlich umfasst der Algorithmus zur Modellierung einer vollständigen Proteinstruktur vier Schritte. Das Rohmodell des Proteins entsteht aus dem Aufsetzen der Ausgangssequenz auf eine geeignete Vorlage. Dem Rohmodell, das nur aus den C α -Atomen des Proteinrückgrats besteht, werden in den Kernregionen Seitenketten hinzugefügt. Anschließend werden die Schleifen modelliert. Zum Schluss werden alle Seitenketten modelliert und optimiert.

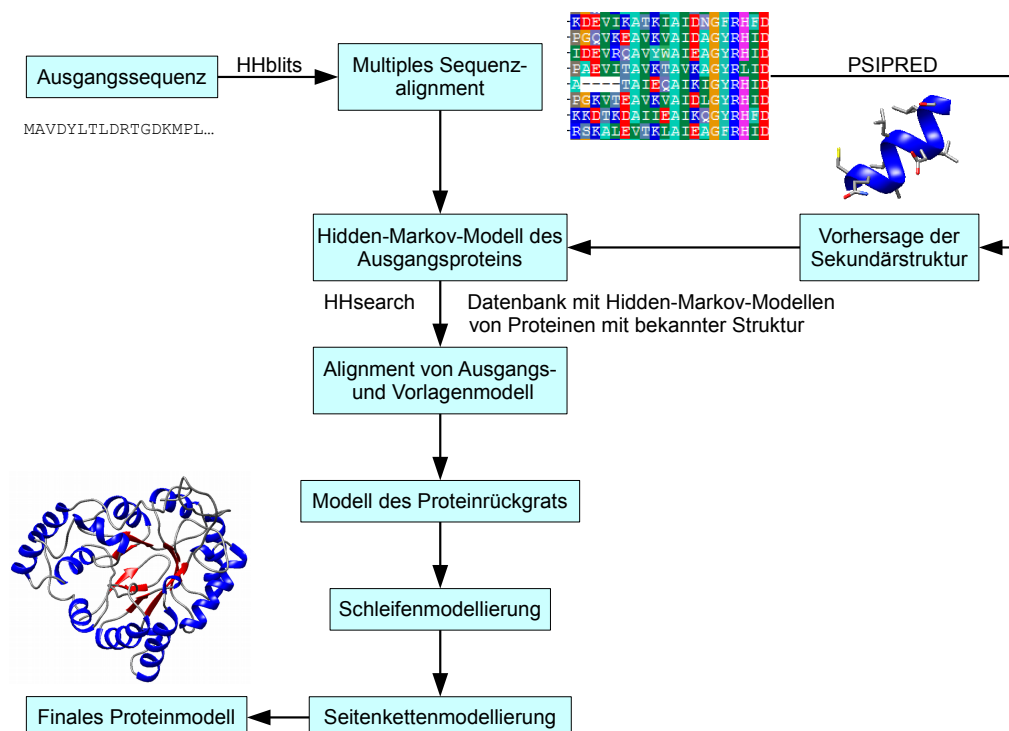


Abbildung 2.2: Ablaufplan der normalen Modellierung mit einer Vorlagenstruktur in Phyre2 am Beispiel des TSP1-Proteins CEP13671 aus *Parasitella parasitica*. (Abbildung nach Kelley et al. (2015))

Als erstes wird ein evolutionäres Profil der Ausgangssequenz erzeugt, das die Aminosäurepräferenz an jeder Position des Proteins darstellt. Dafür wird eine gewisse Anzahl an homologen und dennoch unterschiedlichen Sequenzen benötigt. Diese werden mit Sequenz-Profil-Abgleich gefunden. Hierfür werden HHblits (Remmert et al., 2012) und eine Datenbank, in der die paarweise Sequenzidentität aller Sequenzen unter 20 % liegt, verwendet. Das Ergebnis ist ein Sequenzprofil der Ausgangssequenz. Das entstandene multiple Sequenzalignment wird für die Vorhersage der Sekundärstruktur mit PSIPRED (Jones, 1999) verwendet. Alignment und Strukturvorhersage werden in einem Hidden-Markov-Modell (HMM) kombiniert, mit der im nächsten Schritt eine Faltungsbibliothek, eine Datenbank aus HMMs von Proteinen mit bekannter Struktur, durchsucht wird. Dafür wird HHsearch (Söding, 2005) verwendet. Die besten Alignments dieser Suche werden benutzt, um Rohmodelle des Proteinrückgrats zu berechnen. Diese beinhalten oft Insertionen und Deletionen und noch keine Aminosäureseitenketten.

Insertionen und Deletionen erzeugen Schleifen, die unter Verwendung einer Fragmentbibliothek modelliert werden. Die Fragmentbibliothek enthält Bruchstücke bekannter Proteinstrukturen von 2 – 15 Aminosäuren Länge und wurde mit einer vollständigen Zerlegung der Strukturdatenbank und anschließendem Clustern der Strukturen erzeugt. Eine Lücke im Modell wird von deren Sequenz, der Geometrie der angrenzenden Bereiche und dem Abstand zwischen den Endpunkten charakterisiert. Für Insertionen wird eine Sequenz-Profil-Suche mit der fehlenden eingefügten Sequenz durchgeführt, um Bruchstücke mit ähnlicher Zusammensetzung der Sequenz und Abstand der Endpunkte zu finden. Heraus kommt eine Liste von meist rund 100 potentiell nützlichen Bruchstücken. Auf ähnliche Weise wird für Deletionen die Sequenz innerhalb eines Fensters vor und nach der Deletion verarbeitet. Diese Bruchstücke werden unter Verwendung der Cyclic-Coordinate-Descent-Methode (CCD) (Canutescu und Dunbrack, 2003) an das Rohmodell angepasst. Zum Schluss werden die passendsten Bruchstücke aufgelistet. Abbildung 2.2 zeigt den Ablaufplan der normalen Modellierung mit einer Vorlagenstruktur (PDB: 1MI3) in Phyre2 am Beispiel des TSP1-Proteins CEP13671 aus *Parasitella parasitica*.

Für die intensive Modellierung werden mehrere Vorlagen verwendet. Dieser Modus ermöglicht das Berechnen vollständiger Modelle unter Verwendung verschiedener Vorlagen für unterschiedliche Regionen eines Proteins. Für diesen Schritt wird Poing (Jefferys et al., 2010) verwendet, ein vereinfachter Proteinfaltungssimulator. Mit einer Heuristik wird eine Gruppe von Modellen aus den vorangegangenen Schritten ausgewählt. Diese sollen sowohl die Abdeckung, als auch das Vertrauen in das Modell erhöhen.

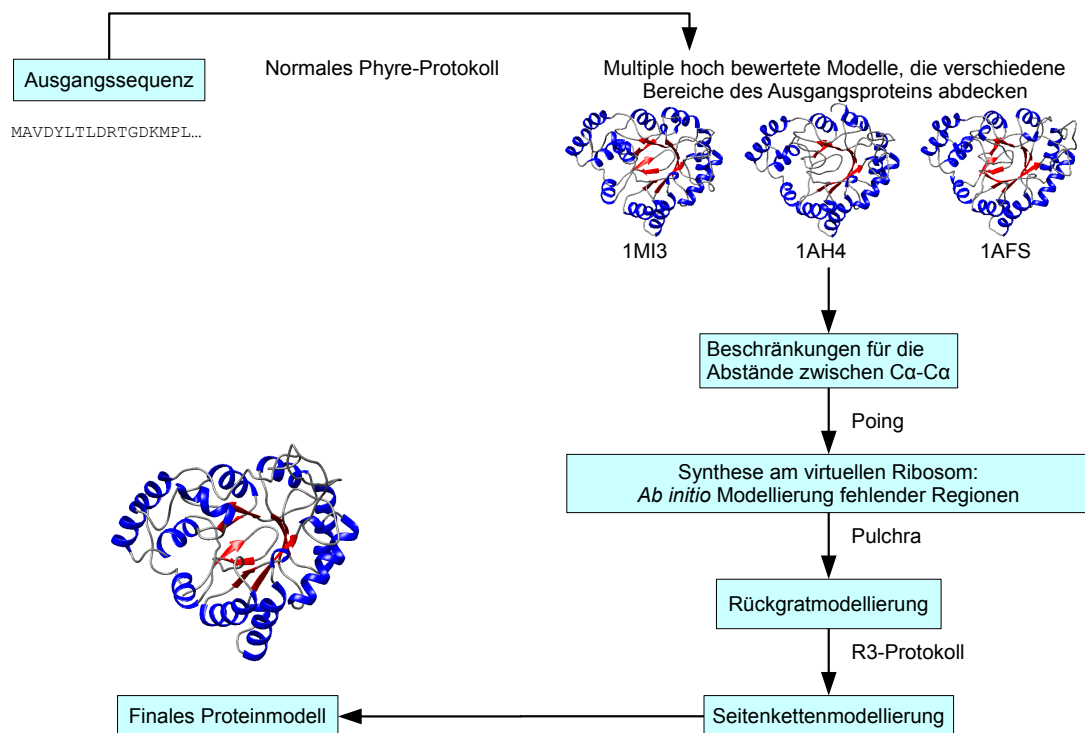


Abbildung 2.3: Ablaufplan der intensiven Modellierung mit mehreren Vorlagenstrukturen in Phyre2 am Beispiel des TSP1-Proteins CEP13671 aus *Parasitella parasitica*. (Abbildung nach Kelley et al. (2015))

Die Ausgangsmodelle liefern Abstandsbedingungen zwischen verschiedenen Aminosäurepaaren. Paarweise Cα–Cα-Abstände werden diesen Modellen entnommen und als lineare starre Federmodelle realisiert. Während das Protein langsam an einem virtuellen Ribosom synthetisiert wird, werden die Beschränkungen in Poing hinzugefügt. Aminosäuren, die nicht von einem der Ausgangsmodelle vorhergesagt werden, lassen sich *ab initio* mit dem Lösungsmittel-Beschuss-Modell (solvent bombardment) in Poing, den vorhergesagten Sekundärstrukturelementen und der Bestrafung sterischer Konflikte erzeugen. Dabei werden zwischen 5 – 100 Modelle synthetisiert, je nach Größe des Proteins. Anschließend werden die Modelle gruppiert, um ein endgültiges Modell auszuwählen. Da dieses Modell nur aus Cα-Atomen besteht, wird das Rückgrat des Proteins mit Pulchra (Rotkiewicz und Skolnick, 2008) rekonstruiert.

Im letzten Schritt werden die Seitenketten der Aminosäuren hinzugefügt, um das endgültige Proteinmodell zu erzeugen. Dafür wird das R3-Protokoll (Xie und Sahinidis, 2006) verwendet. Es bezieht eine schnelle Graph-basierte Technik und eine Seitenkettenbibliothek mit Rotationsisomeren ein, um Seitenketten an der wahrscheinlichsten Position einzufügen und sterische Konflikte zu vermeiden. Abbildung 2.3 zeigt den Ablaufplan der intensiven Modellierung unter Verwendung mehrerer Vorlagenstrukturen (PDB: 1MI3, 1AH4, 1AFS) in Phyre2 am Beispiel des TSP1-Proteins CEP13671 aus *Parasitella parasitica*.

Tabelle 2.1 zeigt besonders häufig gewählte Vorlagenstrukturen für die Modellierung der TSP-Proteine mit Phyre2.

	PDB ID	Proteinname
TSP1	1VP5	2,5-Diketo-D-Glukonsäure-Reduktase
	1Q5M	20 α -Hydroxysteroid-Dehydrogenase
	1AFS, 1J96	3 α -Hydroxysteroid-Dehydrogenase
	1HQT, 2ALR	Aldehyd-Reduktase
	1FRB, 3F7J, 3H7U, 3WCZ	Aldo-Keto-Reduktase
	1AH4, 1US0	Aldose-Reduktase
	1S1P	Prostaglandin-D2-11-Keto-Reduktase
	1VBJ, 4FZI	Prostaglandin-F-Synthase
	1MI3	Xylose-Reduktase
TSP2	1N5D	20 α -Hydroxysteroid-Dehydrogenase
	2P68, 4Z0T	3-Oxoacyl-[Acyl-Carrier-Protein]-Reduktase
	5EPO	7 α -Hydroxysteroid-Dehydrogenase
	1SNY, 1WMA, 1YO6, 3WXB	Carbonyl-Reduktase
	2CFC	Hydroxypropylethan-Thiosulfonat-Dehydrogenase
	3TOX	Kurzkettige Dehydrogenase
	1H5Q	Mannitol-Dehydrogenase
	4IQG, 5IDW	Oxidoreduktase
	3O26	Salutaridin-Reduktase
TSP3	2BIW	Apocarotinoid spaltende Oxygenase
	3FSN	Retinoid-Isomerohydrolase
	3NPE	9-cis-Epoxycarotinoid-Dioxigenase

Tabelle 2.1: Die am häufigsten ausgewählten Vorlagenstrukturen der Homologie-Modellierung mit Phyre2 von TSP1-, TSP2- und TSP3-Proteinen.

Eins-zu-eins-Abbildung (one-to-one threading) ist eine spezielle Methode, die es ermöglicht, auf die automatische Suche nach einer geeigneten Vorlage zu verzichten und diese selbst vorzugeben. Das macht nicht nur Sinn für Anfragen, denen eine noch unveröffentlichte Struktur zugrunde liegt, sondern auch für weiterführende Analysen eines bereits erstellten Proteinmodells. Für die Ausgangssequenz und die Vorlage werden Hidden-Markov-Modelle berechnet und wie bereits beschrieben verarbeitet. Diese Technik wurde für die Modellierung der *in silico* Mutanten von TSP1 aus *Parasitella parasitica* angewendet (siehe Abschnitt 3.5). Als Vorlage wurde die Struktur des bereits modellierten Wildtyp-TSP1-Proteins vorgegeben.

Abschnitt	Beschreibung
Title	Zusammenfassende Beschreibung
Remark	Anmerkungen, wie Autoren, Veröffentlichungen, experimentelle Details
Primary structure	Beschreibung der Primärstruktur
Secondary structure	Beschreibung der Sekundärstruktur
Heterogen	Beschreibung von Nicht-Standardgruppen
Connectivity annotation	Anmerkungen zu speziellen chemischen Bindungen
Miscellaneous features	Anmerkungen zu verschiedenen Besonderheiten
Crystallographic	Beschreibung der kristallografischen Struktur
Coordinate transformation	Beschreibung der Koordinatentransformation
Coordinate	Beschreibung der einzelnen Atome
Connectivity	Beschreibung der chemischen Bindungen
Bookkeeping	Abschließende Bemerkungen

Tabelle 2.2: Abschnitte der Beschreibung einer Proteinstruktur im PDB-Format.

Das endgültige Proteinmodell wird im PDB-Format gespeichert (ftp://ftp.wwpdb.org/pub/pdb/doc/format_descriptions/Format_v33_A4.pdf). Das ermöglicht die direkte Verwendung der vorhergesagten Struktur für weiterführende Simulationen. Tabelle 2.2 zeigt, aus welchen Teilen eine PDB-Datei aufgebaut ist. Jede Zeile beginnt mit einem Schlüsselwort, wie zum Beispiel *Title*, und der entsprechenden Beschreibung. In Abschnitt 2.6 wird der Aufbau der in dieser Arbeit erstellten PDB-Dateien erläutert (Abbildung 2.5, Tabelle 2.3).

2.5 Auswertung der modellierten Proteinstrukturen

ProSA-web (Wiederstein und Sippl, 2007) wurde verwendet, um die Qualität der modellierten Proteinstrukturen zu bewerten. Die Anwendung erkennt Fehler in dreidimensionalen Proteinstrukturen. Mit Hilfe eines Distanz-basierten Paar-Potentials (Sippl, 1990; Sippl, 1995) und eines Potentials, das die Lösemittelzugänglichkeit der Aminosäurereste im Protein beschreibt (Sippl, 1993a) (siehe Abschnitt 1.3.3), wird die Energie des Proteins berechnet. Von diesen Energien werden zwei Eigenschaften der Proteinstruktur abgeleitet, der Z-Wert und die lokale Energie der Aminosäurereste. Der Z-Wert beschreibt die globale Qualität des Proteinmodells und misst die Abweichung der Gesamtenergie der Struktur im Bezug auf eine Energieverteilung, die von zufälligen Verteilungen abgeleitet wurde (Sippl, 1993b; Sippl, 1995). Der Z-Wert darf für gute Modelle nicht außerhalb des für native Proteine charakteristischen Bereichs liegen.

In einer Grafik werden die lokalen Qualitätswerte als Energiefunktion für jede Position innerhalb der Aminosäuresequenz dargestellt. Abbildung 2.4 zeigt die Ausgabe von ProSA-web für ein TSP1-Protein aus *Parasitella parasitica*.

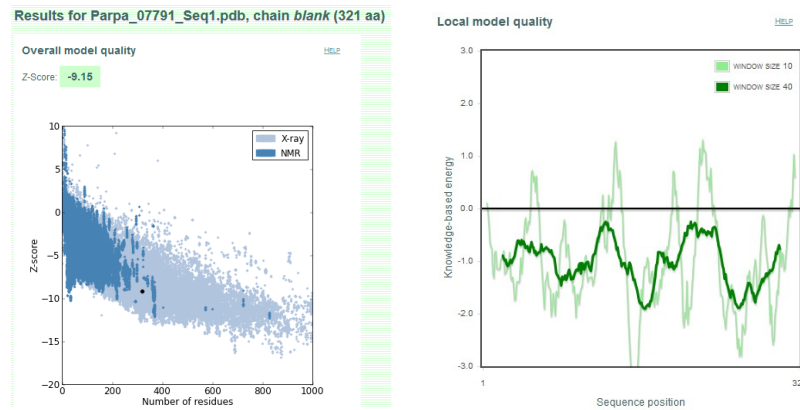


Abbildung 2.4: Ausgabe von ProSA-web für das TSP1-Protein PARPA_07791 aus *Parasitella parasitica*. Der Z-Wert liegt innerhalb der für native Proteine ermittelten Werte. Die lokalen Qualitätswerte zeigen bei geeigneter Fenstergröße (40) keine fehlerhaften Bereiche.

Positive Energiewerte zeigen problematische Abschnitte eines Modells an.

UCSF Chimera (Pettersen et al., 2004) wurde verwendet, um die modellierten Proteinstrukturen, Docking-Ergebnisse, Dimere und Trajektorien zu visualisieren und miteinander zu vergleichen. Die Abbildungen der Proteinstrukturen und der Strukturalignments wurden mit UCSF Chimera erzeugt.

2.6 Protein-Ligand-Docking mit AutoDock Vina

Strukturelle und energetische Eigenschaften von Ligand und Rezeptor bestimmen über die Stabilität der Bindung eines Liganden an ein Protein (Rezeptor). Eine wichtige thermodynamische Größe, die eine Aussage darüber gibt, ob ein Ligand bindet oder nicht, ist das Gibbs-Potential (G). Ist die Änderung des Gibbs-Potentials bei der Bindung negativ ($\Delta G < 0$), so findet die Ligandenbindung spontan statt, anderenfalls ($\Delta G > 0$) wird sie nicht spontan stattfinden. Die Differenz des Gibbs-Potentials ΔG wird mit den beiden in Gleichung 2.1 beschriebenen Zusammenhängen dargestellt.

$$\Delta G = -RT \ln K_i = \Delta H - T \Delta S \quad (2.1)$$

Wie in Gleichung 2.1 gezeigt, wird eine Ligandenbindung sowohl von enthalpischen (ΔH), als auch von entropischen (ΔS) Kräften beeinflusst.

Die Bindungsaffinität eines Liganden wird von der experimentell bestimmbaren Gleichgewichtskonstante K_i beschrieben. K_i gibt an, welcher Anteil des Liganden im Mittel an das Protein gebunden ist und wird von Gleichung 2.2 definiert.

$$K_i = \frac{[Ligand] \times [Protein]}{[Ligand \times Protein]} \quad (2.2)$$

Typische Werte der Bindungskonstante K_i für die Ligandenbindung an ein Protein bewegen sich im Bereich von $10^{-12} - 10^{-2}$ mol/l (M). Dies entspricht einer Änderung des

Gibbs-Potentials von $-70 - -10 \text{ kJ mol}^{-1}$, wobei ein K_i von 10^{-9} M ein guter Wert für einen Wirkstoff ist. Es gilt, je kleiner der K_i -Wert, desto stärker die Bindung.

Mittlerweile sind die dreidimensionalen Strukturen vieler Protein-Ligand-Komplexe und deren Bindungsaffinitäten bekannt. Aus diesen Daten lassen sich einige allgemein gültige Aussagen für die hoch affine Bindung niedermolekularer Liganden und ihrer makromolekularen Rezeptoren treffen:

1. Zwischen Rezeptor und Ligand existiert ein hohes Maß an sterischer Komplementarität. Diese Beobachtung entspricht dem von Emil Fischer formulierten „Schlüssel-Schloss-Prinzip“ (Fischer, 1894).
2. Es gibt eine sehr hohe Übereinstimmung von Oberflächeneigenschaften in Rezeptor und Ligand. Lipophile Teile des Liganden befinden sich meist im Kontakt mit lipophilen Teilen des Rezeptors. Polare Gruppen des Liganden wirken im Wechsel mit geeigneten polaren Gruppen des Rezeptors.
3. Der Ligand bindet meist in einer energetisch günstigen Konformation.
4. Abstoßende Wechselwirkungen sind bei einer spezifischen Bindung äußerst selten.

Strukturdaten von nieder affinen Protein-Ligand-Komplexen, die teilweise aus abstoßenden Wechselwirkungen zwischen Ligand und Rezeptor resultieren, sind sehr selten, da sie zum einen für die meisten Strukturbiologen weniger interessant und zum anderen viel schwerer zu erhalten sind. Diese Daten sind jedoch unverzichtbar für die Entwicklung von Kraftfeldern und Bewertungsfunktionen als Grundlage für ein computergestütztes strukturbasiertes Liganden-Design. Im Wesentlichen lassen sich folgende Aussagen machen:

1. Polare Gruppen an der Grenzfläche zwischen Rezeptor und Ligand, denen ein entsprechender Partner fehlt, verringern die Bindungsaffinität erheblich.
2. Führt die Bindung des Liganden zu einer Verringerung der Gesamtzahl an Wasserstoffbrücken, verringert das die Bindungsaffinität.
3. Eine nicht optimale sterische Korrelation zwischen Rezeptor und Ligand führt zu Löchern und Blasen im lipophilen Teil der Rezeptor-Ligand-Grenzfläche und resultiert in schlechteren Bindungsaffinitäten.

Entropische Einflüsse haben zum Beispiel die Wassermoleküle des Lösungsmittels, die an einer hydrophoben Bindungstasche stark geordnet sind und nach der Bindung eines passenden Liganden an das Lösungsmittel abgegeben werden. Da Wassermoleküle in Lösung deutlich ungeordneter sind, führt die Bindung des Liganden zu einem Entropiegewinn. Dieser Vorgang wird als der hydrophobe Effekt bezeichnet. Der hydrophobe Effekt

hat allerdings noch einen zweiten enthalpischen Beitrag, da die Wassermoleküle in einer hydrophoben Bindungstasche weniger Wasserstoffbrücken ausbilden, als in der freien Lösung. Zwischen Ligand und Protein gibt es drei wichtige nicht kovalente Wechselwirkungen, die zu einem Enthalpiegewinn beitragen:

1. **Elektrostatische Wechselwirkungen:** Die Coulomb-Wechselwirkung resultiert aus der Kraft zwischen entgegengesetzten Ladungen und ist stark von dem Abstand der Ladungen abhängig. Eine wichtige Eigenschaft der elektrostatischen Wechselwirkung ist, dass sie ungerichtet ist.
2. **Wasserstoffbrückenbindungen:** Wasserstoffbrücken (H-Brücken) existieren sowohl zwischen geladenen, als auch ungeladenen Molekülen. Bei einer H-Brücke teilen sich zwei Atome ein Wasserstoffatom. Es gibt stets einen H-Donor, an den der Wasserstoff kovalent gebunden ist, und einen H-Akzeptor. Der H-Donor muss eine so hohe Elektronegativität besitzen, dass der Wasserstoff partiell positiv geladen ist und der H-Akzeptor muss ein freies Elektronenpaar besitzen. Charakteristisch sind Abstände von 2,5 – 3,2 Å, sowie Winkel im Bereich von 130 – 180°. H-Brücken sind im Gegensatz zu Coulomb-Wechselwirkungen gerichtet. Ungünstige Winkel oder Abstände vermindern die Energie stark.
3. **Van-der-Waals-Wechselwirkungen:** Van-der-Waals-Wechselwirkungen sind unspezifische, schwache Wechselwirkungen, die erst zum Tragen kommen, wenn eine größere Fläche beteiligt ist. Die Stärke dieser hydrophoben Wechselwirkungen wird über die lipophile Fläche beschrieben, die bei der Bindung begraben wird. Der Energiegehalt liegt im Bereich von 100 – 200 Jmol⁻¹ Å⁻². Van-der-Waals-Wechselwirkungen kommen zustande, indem eine zeitlich variierende Elektronenverteilung um einen Kern die Elektronenverteilung der Nachbaratome beeinflusst und so eine Anziehung entsteht.

Die Bewertungsfunktionen, die für Protein-Ligand-Docking angewendet werden, sind in der Lage, freie Bindungsenergien im Bereich von 7 – 10 kJ/mol vorherzusagen. Sie werden in drei Hauptklassen unterteilt: Kraftfeld-basierte, empirische und wissensbasierte Bewertungsfunktionen (Sousa et al., 2006).

Um Aussagen über die Substratspezifität der einzelnen TSP-Proteine zu treffen, wurde Protein-Ligand-Docking mit UCSF Chimera (Pettersen et al., 2004) und AutoDock Vina (Trott und Olson, 2010) durchgeführt.

AutoDock Vina lässt sich sowohl mit einem starren, als auch einem beweglichen Rezeptor starten. Zur Vereinfachung wurde hier von einem starren Rezeptor ausgegangen. Dies ist natürlich physikalisch nicht richtig, denn ein Protein ist nicht starr, sondern befindet sich in Lösung in dynamischer Bewegung. Die Bindung eines Liganden führt zu

einer strukturellen Veränderung im Rezeptormolekül (Induced-Fit). Eine genaue Modellierung dieser Vorgänge ist jedoch mit sehr hohen Rechenzeiten verbunden. Die Annahme eines starren Rezeptors reduziert die Rechenzeit erheblich und führt in vielen Fällen zu zufriedenstellenden Ergebnissen. Während einige Docking-Programme in der Lage sind, Seitenketten in der Bindungstasche geringfügig anzupassen, zum Beispiel um Wasserstoffbrücken zu optimieren, wird die dynamische Natur des Bindungsvorgangs nur in Moleküldynamik-Simulationen explizit modelliert.

```

-> Modell 1 enthält die vorhergesagte Proteinstruktur zusammen
mit dem gebundenen Cofaktor NADP.
MODEL          1      1
ATOM           1  N   MET           1      22.312 -20.069  25.663  1.00  0.00      N
:
ATOM          5198  HH   TYR          321      28.507  22.601  24.302  1.00  0.00      H
TER           5199      TYR          321
HETATM        5200  P1   NAP           1      25.342  -8.182  38.993  1.00  0.00      P
:
HETATM        5257  H17  NAP           1      31.480  -4.744  43.938  1.00  0.00      H
ENDMDL
-> Modell 2 enthält die vorhergesagte Proteinstruktur zusammen
mit dem gebundenen Trisporoids substrat.
MODEL          2
ATOM           1  N   MET           1      22.313 -20.069  25.663  1.00  0.00      N
:
ATOM          5198  HH   TYR          321      28.893  22.914  25.788  1.00  0.00      H
TER           5199      TYR          321
HETATM        5200  C   TB1           0      39.014   1.787  35.079  0.00  0.00      C
:
HETATM        5223  H   TB1           0      35.998   1.021  35.572  0.00  0.00      H
ENDMDL
12345678901234567890123456789012345678901234567890123456789012345678901234567890

```

Abbildung 2.5: Beispiel für eine Datei im PDB-Format, die ein TSP1-Protein im Komplex mit dem Cofaktor NADP (NAP) und einem Trisporoid (TB1) zeigt. Die letzte Zeile gibt die Spalten innerhalb der ATOM-Zeilen an, deren Einträge in Tabelle 2.3 erklärt werden.

Die Strukturinformationen von Rezeptor und Ligand wurden AutoDock Vina in Form von PDB-Dateien übergeben. Abbildung 2.5 zeigt einen Ausschnitt aus der PDB-Datei eines TSP1-Proteins nach Proteinstrukturvorhersage und Protein-Ligand-Docking. Modell 1 beinhaltet das beste Modell des Protein-Ligand-Dockings für das Cosubstrat NADP, abgekürzt mit NAP. Modell 2 beinhaltet das beste Modell des Protein-Ligand-Dockings für das Substrat 4-Dihydromethyltrisporat der B-Serie, abgekürzt mit TB1. Die letzte Zeile gibt die Spalten innerhalb der ATOM-Zeilen an, deren Einträge in Tabelle 2.3 erklärt werden. Die PDB-Dateien der TSP-Proteine beinhalten nur einen Teil dieser Informationen. Auf eine Bezeichnung der Proteinketten wurde verzichtet, da es sich um Monomere handelt.

Einträge in der RCSB Protein Data Bank (PDB) (Rose et al., 2015) stammen in der Regel entweder von Strukturbestimmungen mit Röntgenkristallographie oder Kernspinresonanzspektroskopie. Da die Proteinstrukturen von TSP-Proteinen nicht bekannt sind, fehlt die für einen dynamischen Rezeptor notwendige Strukturinformation. Solche spezifischen Strukturdetails sind meist Artefakte der Homologie-Modellierung.

Spalten	Datentyp	Beschreibung
1 - 6	String	Schlüsselwort der Zeile
7 - 11	Integer	Fortlaufende Nummer des Atoms
13 - 16	String	Name des Atoms
17	Char	Alternativer Positionsbezeichner
18 - 20	String	Name des Aminosäurerests
22	Char	Bezeichner der Proteinkette
23 - 26	Integer	Sequenznummer der Aminosäure
27	Char	Code zum Einfügen eines Aminosäurerests
31 - 38	Real	X Koordinate in Angström
39 - 46	Real	Y Koordinate in Angström
47 - 54	Real	Z Koordinate in Angström
55 - 60	Real	Belegungsfaktor
61 - 66	Real	Temperaturfaktor
73 - 76	String	Segmentbezeichner
77 - 78	String	Elementsymbol
79 - 80	String	Ladung des Atoms

Tabelle 2.3: Aufbau einer ATOM-Zeile zur Beschreibung von Atomkoordinaten in einer PDB-Datei.

AutoDock Vina verwendet zur Berechnung der Interaktionsenergie eine empirische Funktion zur Bestimmung der freien Bindungsenergie. Die allgemeine funktionelle Formel des konformationsabhängigen Teils der Bewertungsfunktion in AutoDock Vina (Trott und Olson, 2010) ist in Gleichung 2.3 gezeigt.

$$c = \sum_{i < j} f_{t_i t_j}(r_{ij}) \quad (2.3)$$

Dabei erfasst die Summe alle Atompaare, die miteinander in Verbindung stehen. Jedem Atom i werden ein Typ t_i und eine symmetrische Gruppe an Interaktionsfunktionen $f_{t_i t_j}$ zugewiesen, die über die Atomabstände r_{ij} berechnet werden. Dieser Wert ist die Summe der inter- und intramolekularen Beiträge (Gleichung 2.4).

$$c = c_{\text{inter}} + c_{\text{intra}} \quad (2.4)$$

Der Optimierungsalgorithmus sucht das globale Minimum von c und andere Konformationen mit niedriger Bewertung, die er anschließend ordnet. Die vorhergesagte freie Bindungsenergie wird mit dem intermolekularen Anteil der am niedrigsten bewerteten Konformationen (bezeichnet mit 1) berechnet (Gleichung 2.5). Die Funktion g ist eine beliebige streng ansteigende, möglicherweise nicht-lineare Funktion.

$$s_1 = g(c_1 - c_{\text{intra}1}) = g(c_{\text{inter}1}) \quad (2.5)$$

In der Ausgabe werden weiteren gering bewerteten Konformationen ebenfalls Werte für s zugewiesen (Gleichung 2.6).

$$s_i = g(c_i - c_{\text{intra}}) \quad (2.6)$$

Die Herleitung der Bewertungsfunktion von AutoDock Vina kombiniert die Vorteile wissensbasierter Potentiale und empirischer Bewertungsfunktionen. Es werden empirische Informationen von Konformationspräferenzen der Rezeptor-Liganden-Komplexe und experimentellen Messungen extrahiert.

Wasserstoffatome werden ausschließlich für die Zuweisung von Atomtypen berücksichtigt. Die Interaktionsfunktionen $f_{t_i t_j}$ werden in Relation zu den Oberflächenabständen $d_{ij} = r_{ij} - R_{t_i} - R_{t_j}$ definiert (Gleichung 2.7). R_t ist der van-der-Waals-Radius von Atomtyp t . $h_{t_i t_j}$ ist eine gewichtete Summe sterischer Interaktionen, die für alle Atompaare identisch sind. Dabei handelt es sich um hydrophobe Interaktionen zwischen hydrophoben Atomen und Wasserstoffbindung.

$$f_{t_i t_j}(r_{ij}) \equiv h_{t_i t_j}(d_{ij}) \quad (2.7)$$

Gewicht	Term
-0,0356	gauss ₁
-0,00516	gauss ₂
0,840	Abstoßung
-0,0351	Hydrophob
-0,587	Wasserstoffbindung
0,0585	N_{rot}

Tabelle 2.4: Gewichte und Terme, die in der Bewertungsfunktion von AutoDock Vina eingesetzt werden. (Tabelle entnommen aus Trott und Olson (2010))

Tabelle 2.4 zeigt die eingesetzten Parameter. Die sterischen Terme werden in den Gleichungen 2.8 – 2.10 dargestellt.

$$\text{gauss}_1(d) = e^{-(d/0,5 \text{ \AA})^2} \quad (2.8)$$

$$\text{gauss}_2(d) = e^{-((d-3 \text{ \AA})/2 \text{ \AA})^2} \quad (2.9)$$

$$\text{repulsion}(d) = \begin{cases} d^2, & \text{if } d < 0 \\ 0, & \text{if } d \geq 0 \end{cases} \quad (2.10)$$

Der hydrophobe Term entspricht 1, wenn $d < 0,5 \text{ \AA}$ und 0, wenn $d > 1,5 \text{ \AA}$. Zwischen diesen beiden Werten wird der Term linear interpoliert. Der Term für die Wasserstoffbindung entspricht 1, wenn $d < -0,7 \text{ \AA}$ und 0, wenn $d > 0$. Zwischen diesen beiden Werten wird der Term ähnlich dem ersten Term linear interpoliert. Metalle werden als Wasserstoffbindungsdonoren behandelt. Alle Interaktionsfunktionen f_{itj} werden bei $r_{ij} = 8 \text{ \AA}$ beschnitten.

Die konformationsunabhängige Funktion g ist in Gleichung 2.11 gezeigt. N_{rot} ist die Anzahl an aktiven rotierbaren Bindungen zwischen schweren Atomen im Liganden. w ist das entsprechende Gewicht.

$$g(c_{\text{inter}}) = \frac{c_{\text{inter}}}{1 + wN_{\text{rot}}} \quad (2.11)$$

Als Optimierungsalgorithmus ist in AutoDock Vina eine iterierte lokale Suche mit globaler Optimierung implementiert, ähnlich dem in Abagyan et al. (1994) beschriebenen Algorithmus. In diesem Algorithmus bestehen die einzelnen Schritte aus einer Mutation der Konformation des Liganden und dessen lokaler Optimierung. Die einzelnen Schritte werden nach dem Metropolis-Kriterium angenommen oder verworfen. Für die lokale Optimierung wird die Methode nach Broyden-Fletcher-Goldfarb-Shanno (BFGS) (Nocedal und Wright, 1999) angewendet. Dabei handelt es sich um eine effiziente quasi-Newton-Methode. Diese verwendet nicht nur den Wert der Bewertungsfunktion, sondern auch deren Gradienten. So zum Beispiel die Ableitung der Bewertungsfunktion in Abhängigkeit zu ihren Argumenten, die aus der Lage und Orientierung des Liganden, sowie den Werten der Torsionswinkel für die aktiven rotierbaren Bindungen im Liganden und in flexiblen Aminosäureresten bestehen. Die Anzahl an Schritten eines Programmlaufs wird von der Komplexität des Problems bestimmt. Mehrere Programmläufe werden mit zufällig gewählten Konformationen gestartet. AutoDock Vina verwendet Multithreading, um diese Programmläufe auf mehrkernigen CPUs parallel zu bearbeiten. Der Optimierungsalgorithmus kombiniert die Ergebnisse aus den gefundenen Minima der getrennten Programmläufe und verwendet sie für die Verbesserung der Struktur und Clusterbildung.

Abbildung 2.6 zeigt die für das Protein-Ligand-Docking benötigten Parameter und mögliche Schritte der Vorverarbeitung. Wassermoleküle und Ketten, die aus untypischen Aminosäuren bestehen, werden ignoriert. Der Suchraum wurde mit einer Größe von rund $20 \text{ \AA} \times 18 \text{ \AA} \times 22 \text{ \AA}$ in der Nähe des aktiven Zentrums über die Trisporoidbindeta-sche definiert. Phyre2 (Kelley et al., 2015) liefert PDB-Dateien, die direkt für Docking-Experimente verwendbar sind und ermöglicht es, Proteine ohne bekannte Proteinstruktur zu untersuchen. Die PDBQT-Datei des Makromoleküls wurde aus dieser PDB-Datei erzeugt. Auch für den Liganden, in diesem Beispiel 4-Dihydromethyltrispomat, benötigt man eine PDBQT-Datei. In der Vorbereitung werden dem Proteinmodell Wasserstoffe hinzugefügt, Ladungen verbunden und nicht-polare Wasserstoffe entfernt.

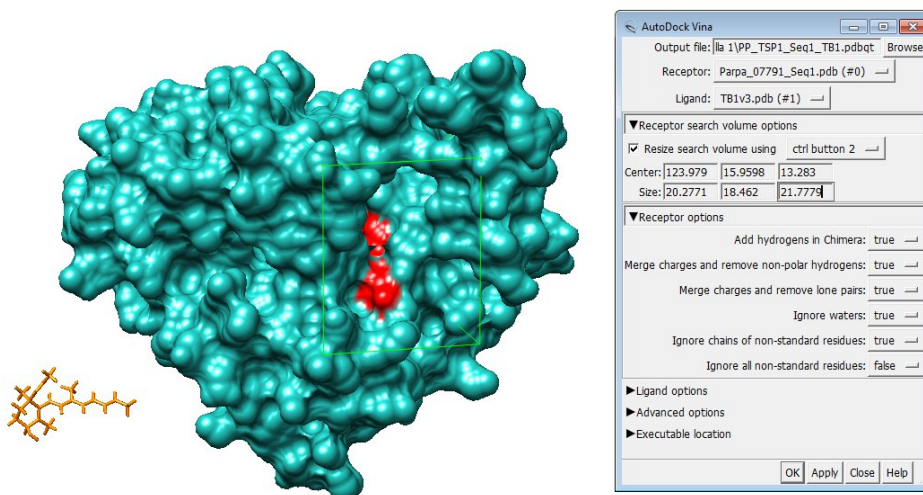


Abbildung 2.6: Eingabemaske der für AutoDock Vina benötigten Parameter in UCSF Chimera. Neben den Strukturen von Rezeptor und Ligand müssen die Größe und Lage des Suchraums (grüne Box) und die Schritte der Vorverarbeitung definiert werden. Abgebildet ist ein TSP1-Protein aus *Parasitella parasitica* (CEP13671) und der Ligand 4-Dihydromethyltrisporat (oranges Molekül). Das aktive Zentrum des Enzyms ist rot markiert.

Anschließend werden dem Modell AD4-Typen zugewiesen. Diese unterscheiden eine PDBQT- von einer PDB-Datei. Es handelt sich dabei um eine Beschreibung der einzelnen Atome, die den Bindungszustand mit ein bis zwei Buchstaben codiert. Zum Beispiel steht NA für einen Akzeptor, der aus einem an ein Wasserstoffatom gebundenen Stickstoffatom besteht.

Als Ligand wurden das Cosubstrat NADP und acht verschiedene Trisporoide aus unterschiedlichen Serien verwendet. Diese wurden nach ihrer Funktion und Serie abgekürzt. *TB1* bezeichnet das Substrat von TSP1, 4-Dihydromethyltrisporat, aus der B-Serie. *TE2* steht für das Substrat von TSP2, 4-Dihydrotrisporin, aus der E-Serie. Die dreidimensionale Struktur von NADP stammt aus der RCSB Protein Data Bank (PDB) (Rose et al., 2015). Die Struktur wurde als SDF (Structured Data File) (Dalby et al., 1992) heruntergeladen und in eine Datei im PDB-Format umgewandelt, die als Ausgangsdatei für das Protein-Ligand-Docking benötigt wird. Dazu wurde UCSF Chimera (Pettersen et al., 2004) verwendet. Die Docking-Experimente mit dem Cosubstrat NADP wurden mit denselben Parametern gestartet, wie die der Trisporoid-Liganden.

2.7 Protein-Protein-Docking mit ClusPro

Die Bildung von Dimeren aus den TSP-Proteinen von Zygomyceten wurde mit Protein-Protein-Docking modelliert. Dafür wurde ClusPro (Kozakov et al., 2013) verwendet. Der Server benutzt PIPER (Kozakov et al., 2006), ein auf schnellen Fourier-Transformationen (FFT) basierendes Docking-Programm, das paarweise Interaktionspotentiale als Teil sei-

ner Bewertungsfunktion (Gleichung 2.12) einsetzt.

$$E = E_{attr} + w_1 E_{rep} + w_2 E_{elec} + w_3 E_{pair} \quad (2.12)$$

E_{attr} und E_{rep} sind die anziehenden und abstoßenden Beiträge zu der Energie der van-der-Waals-Interaktionen. E_{elec} ist ein Term für die elektrostatische Energie. E_{pair} steht für den Beitrag der Desolvationsenergie. Die Koeffizienten w_1 , w_2 und w_3 definieren das Gewicht der einzelnen Terme. ClusPro generiert vier Modellarten gleichzeitig nach unterschiedlichen Bewertungsstrategien. Die erste ist zwischen allen Wechselwirkungen ausgeglichen. Die zweite bevorzugt elektrostatische Wechselwirkungen. Die dritte bevorzugt hydrophobe Wechselwirkungen. Und die vierte kombiniert van-der-Waals-Interaktionen und elektrostatische Interaktionen. Bei der letzten Variante wird auf die Verwendung der paarweisen Potentiale E_{pair} verzichtet.

Die 1 000 besten von PIPER erzeugten Strukturen für jede der vier Bewertungsstrategien werden zu Clustern zusammengefasst. Dafür wird der paarweise RMSD-Wert (Root Mean Square Deviation) zur Messung der Distanz verwendet. Auf Energie-basiertes Clustern wird verzichtet, da gerade im Bereich niedriger Energieniveaus die Unterschiede zwischen einzelnen Modellen zu gering sind. Bei großen Clustern ist die Wahrscheinlichkeit höher, dass sie native Proteinstrukturen enthalten. ClusPro gibt die zentralen Modelle der zehn größten Cluster aus.

Mit Hilfe von ClusPro wurde die Bildung von Dimeren aus TSP-Proteinen modelliert. Es wurde nach Unterschieden zwischen Homo- und Heterodimeren gesucht. Besonderes Interesse galt Heterodimeren, die aus gemischten TSP1-Proteinen aus *Absidia glauca* und *Parasitella parasitica* bestanden.

Jedes Sequenzpaar wurde zweimal gestartet, da die resultierenden Dimere keine Symmetrie aufweisen. Das Ergebnis ist von der Art des Proteins in der Modellierung abhängig. Ein Protein wird als Rezeptor oder als Ligand definiert. Ein Dimer, bei dem Sequenz 1 der Rezeptor und Sequenz 2 der Ligand ist, sieht anders aus, als ein Dimer, bei dem Sequenz 2 der Rezeptor und Sequenz 1 der Ligand ist. Für die weiteren Analysen wurden die Modelle verwendet, die hydrophobe Interaktionen bevorzugen.

2.8 Modellierung lokaler Proteindynamik mit PELE

Um ein gewisses Maß an Dynamik in die Modelle einfließen zu lassen, wurden lokale Bewegungen der Monomere und Dimere mit Hilfe von Protein Energy Landscape Exploration (PELE) (Madadkar-Sobhani und Guallar, 2013) simuliert. Das Programm läuft auf einem Computercluster mit 144 CPUs. Für die Simulation der lokalen Dynamiken innerhalb eines Proteins wird die Bewegung des Proteins in rund 100 Schritten unter Verwendung anisotropischer Netzwerkmodelle (ANM, anisotropic network model) (Atilgan et al., 2001) realisiert. Diese sind an eine Optimierung der Seitenketten und einen

abschließenden Minimierungsschritt gebunden. Die zufällige Bewegung des Proteinrückgrats folgt einer Richtung, die für sechs Monte-Carlo-PELE-Schritte ausgewählt wird. Nach diesen sechs Schritten wechselt die Bewegung in eine neue zufällig gewählte Richtung. Die Vorhersage der Seitenketten beinhaltet N (Standardwert $N = 25$) Seitenketten, die während der Bewegung angeregt werden. Die abschließende Minimierung beinhaltet eine Einschränkung von 10 kcal/mol für die C_{α} -Atome, die von der anfänglichen Störung des Proteins abgeleitet wird.

Es wird ein Graph ausgegeben, der den Verlauf der freien Energie des Proteinmodells über die 100 Schritte darstellt (Abbildung 2.7). Die Energieniveaus der ersten beiden Modelle, das heißt die Energie des Ausgangsmodells und die Energie nach dem ersten Minimierungsschritt, sowie des Modells mit der niedrigsten Energie wurden untereinander verglichen. Die PDB-Dateien der Trajektorien befinden sich auf dem zu dieser Arbeit gehörenden Datenträger.

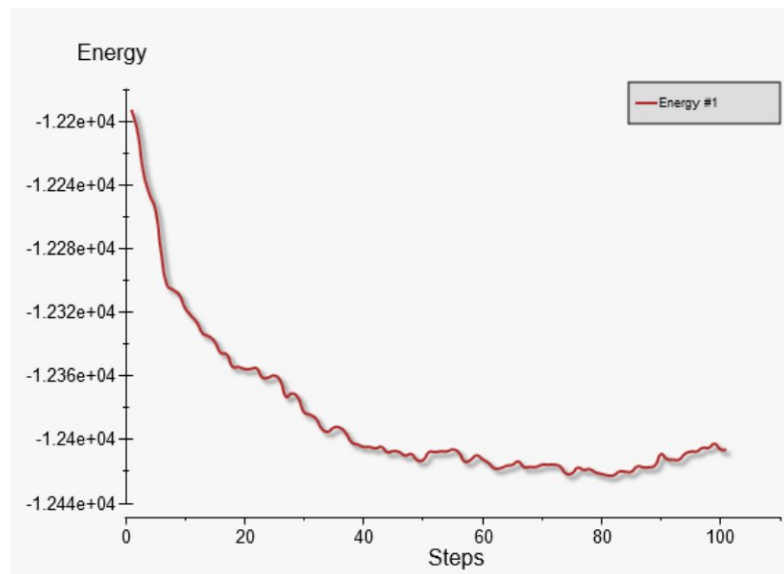


Abbildung 2.7: Energieverlauf der Modellierung lokaler Proteinbewegung mit PELE (kcal/mol) für ein TSP1-Protein aus *Parasitella parasitica* (CEP13671).

UCSF Chimera (Pettersen et al., 2004) wurde dazu verwendet, die Trajektorien in kleinen Videos zu veranschaulichen (siehe Abschnitt 6.1).

2.9 Programmentwicklung

Im Rahmen dieser Arbeit entstanden einige kleinere Java-Programme, wie zum Beispiel GC-Draw für die ringförmige Darstellung des GC-Gehalts von mitochondrialer DNA (Abbildung 3.6) und mehrere Skripte, die unter anderem die Integration der Genvorhersagen mit AUGUSTUS, deren BLAST-Ergebnissen und identifizierten Proteindomänen zur Annotation von Genfunktionen für die Erstellung der Datenbankeinträge für die beiden sequenzierten Genome von *Absidia glauca* und *Parasitella parasitica* in ENA (European Nucleotide Archive) realisierten. Die Algorithmus-Implementierung für die Detektion von horizontalem Gentransfer zwischen Bakterien und Protozoen entstand wie alle anderen Java-Projekte mit Eclipse IDE 4.5.1 (<https://eclipse.org/mars/>) und Java 8 (<https://www.java.com/de/download/faq/java8.xml>). Ein anderes Programm realisiert eine graphische Benutzeroberfläche, die das Arbeiten mit den neuen Genomen und den Zugriff auf die für TSP1-Proteine erstellten Videos erleichtert.

2.9.1 Detektion von horizontalem Gentransfer: HGT Calculator

Die gezielte Suche nach horizontalem Gentransfer (HGT) unterscheidet sich je nachdem, ob das HGT-Ereignis zwischen Bakterien oder zwischen Prokaryoten und Eukaryoten stattgefunden hat. Es gibt verschiedene Ansätze, um HGT-Ereignisse zu finden. Oft ist nicht klar, welche Methode am besten geeignet ist. Zudem weichen Ergebnisse, die mit unterschiedlichen Programmen entstanden sind, zum Teil deutlich voneinander ab. Deshalb wurden verschiedene Methoden in einer Anwendung kombiniert, um die Sensitivität der Suche zu erhöhen. HGT Calculator analysiert die Sequenzen eines Datensatzes in vier Stufen, sowohl auf Nukleotid- als auch auf Proteinebene. Es werden parametrische und Baum-basierte Verfahren angewendet. Tabelle 2.5 zeigt Eigenschaften, die auf ein HGT-Ereignis hinweisen. Diese Eigenschaften entscheiden über den berechneten HGT-Wert, der die Wahrscheinlichkeit für ein HGT-Ereignis angibt.

Proteinsequenz	BLAST-Ergebnisse
<ul style="list-style-type: none"> → untypische Sequenzlänge → Zuordnung zu einer KEGG-Orthologie → unterschiedliche Proteinunterarten → untypische Proteindomäne → hoher Anteil an Prokaryoten für diese Proteindomäne 	<ul style="list-style-type: none"> → Mehrheit der BLAST-Treffer aus Prokaryoten Der Anteil prokaryotischer Sequenzen wird in drei Gruppen gewertet: 50% - 65%, 65% - 90% oder > 90% → Alle eukaryotischen BLAST-Treffer aus einem Organismus → Alien-Index ≥ 45
Nukleotidsequenz	Phylogenetische Bäume
<ul style="list-style-type: none"> → untypischer GC-Gehalt: im Durchschnitt an erster Codonposition an zweiter Codonposition an dritter Codonposition → niedriger Codon Adaptation Index (CAI) 	<ul style="list-style-type: none"> → Nachbar eines Prokaryoten im einzelnen, kleinen und großen Baum → großer Baum enthält eukaryotischen und prokaryotischen Teilbaum → großer Baum enthält Teilbäume für alle Proteinunterarten

Tabelle 2.5: Untersuchte Eigenschaften von HGT-Kandidaten auf Nukleotid- und Proteinebene. (Tabelle nach Ellenberger, Siegmund et al. (2016))

HGT Calculator wurde an zehn bereits publizierten HGT-Genen aus Trypanosomati- den getestet. Zehn weitere Gene wurden als Negativkontrolle mit ursprünglichen Genen, die nicht das Resultat von horizontalem Gentransfer sind, verwendet. Die Liste der Orga- nismen, die für diesen Datensatz ausgewählt wurden, beinhaltet sieben Protozoen (*Leis- hmania*, *Paramecium*, *Plasmodium*, *Strigomonas*, *Tetrahymena*, *Toxoplasma*, *Trypanoso- ma*), einen Pilz (*Saccharomyces*), eine Pflanze (*Arabidopsis*), ein Metazoon (*Homo*), und Bakterien aus 16 verschiedenen taxonomischen Gruppen (*Aquifex*, *Bacillus*, *Chlorobium*, *Cytophaga*, *Deinococcus*, *Desulfovibrio*, *Escherichia*, *Helicobacter*, *Micrococcus*, *Neis- seria*, *Nostoc*, *Sinorhizobium*, *Streptomyces*, *Synechococcus*, *Thermotoga*, *Vibrio*). Für spätere Untersuchungen wurde die taxonomische Zuordnung aus der UniProt-Datenbank (www.uniprot.org/taxonomy) verwendet. Die Liste der zu berücksichtigenden Orga- nismen kann zu jeder Zeit verändert werden. Abbildung 2.8 zeigt die beiden Fenster des HGT Downloaders, eines Teilprogramms des HGT Calculators, in denen sich eukaryoti- sche und prokaryotische Organismen auswählen lassen.

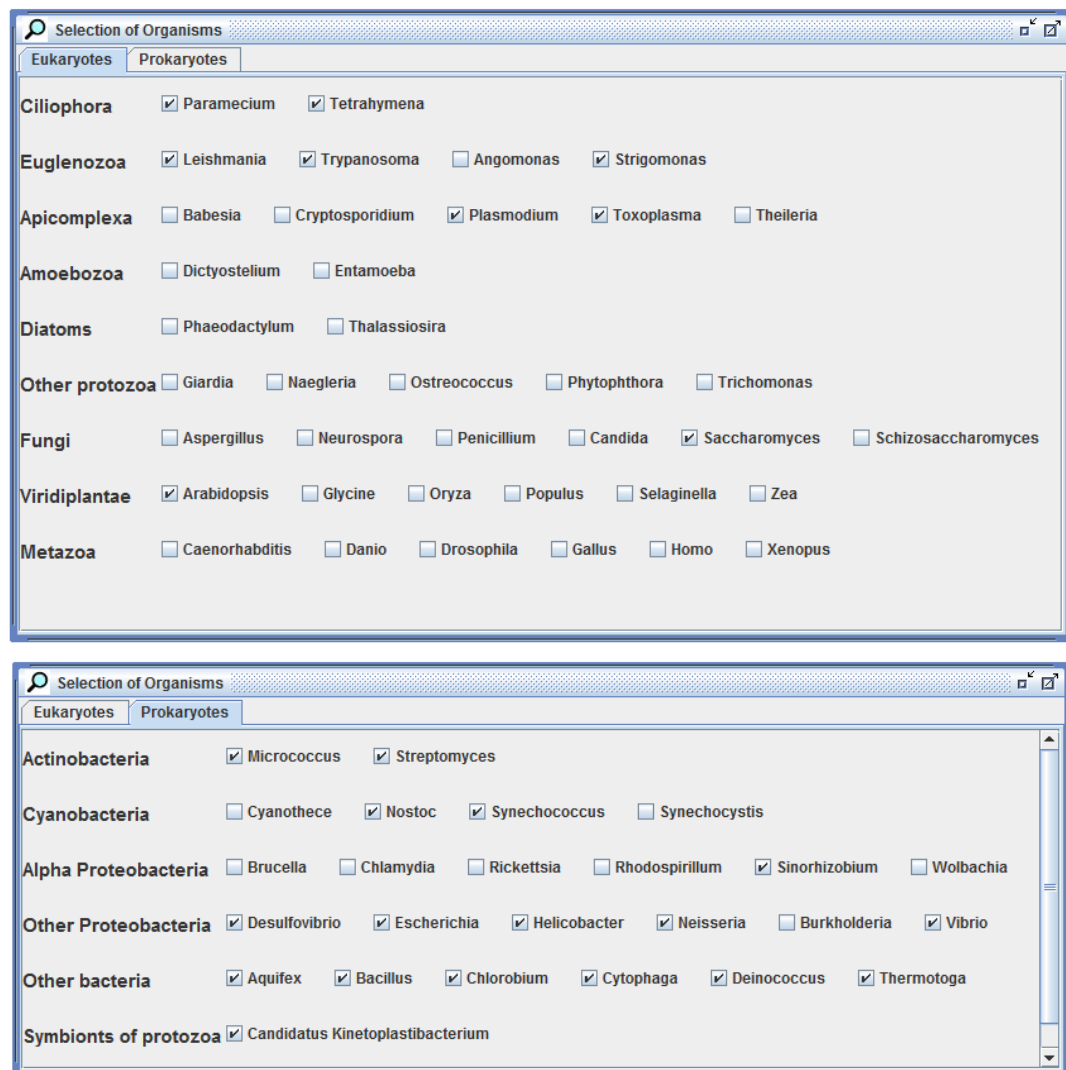


Abbildung 2.8: HGT Downloader: Änderungen in der Auswahl der zu berücksichtigenden Orga- nismen sind jederzeit möglich.

Bevor man nach HGT-Ereignissen suchen kann, braucht es einen geeigneten Datensatz. Für die Sequenzauswahl und Vorverarbeitung wurde der HGT Downloader entwickelt (Abbildung 2.9), eine Eingabemaske für die graphische Benutzeroberfläche des HGT Calculators.

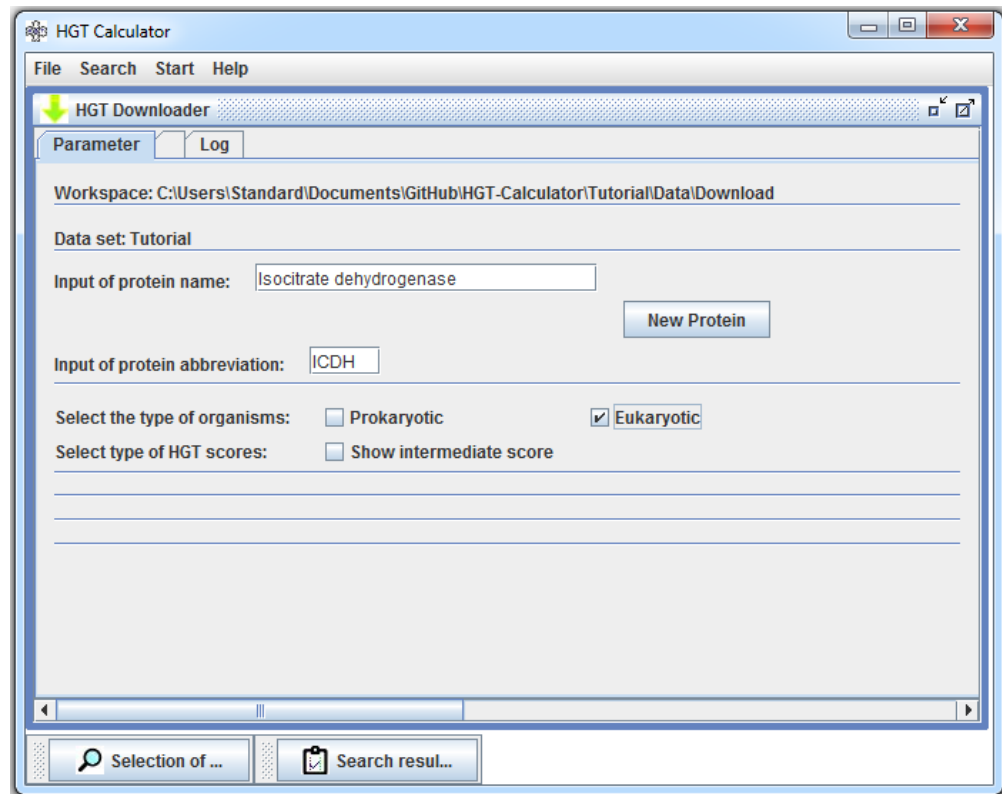


Abbildung 2.9: HGT Downloader unterstützt die Zusammenstellung und Vorverarbeitung von Datensätzen für HGT Calculator.

Der HGT Downloader bietet drei verschiedenen Möglichkeiten, um nach geeigneten Sequenzen für die Zusammenstellung eines individuellen Datensatzes zu suchen: (i) mit einer Identifikationsnummer für eine KEGG-Orthologie (Kanehisa et al., 2012), (ii) mit einer UniProt-Identifikationsnummer (Consortium, 2015) oder (iii) mit einer Textsuche in UniProt. Abbildung 2.10 zeigt die Ausgabe verschiedener Anfragen zur Suche nach Isocitrat-Dehydrogenasen. Die ausgewählten Sequenzen werden nach ihren Proteindomänen sortiert und in einer Liste gespeichert. Es wird getrennt nach prokaryotischen und eukaryotischen Sequenzen gesucht. Die Suche beginnt immer mit den eukaryotischen Sequenzen, da deren Einteilung in Gruppen für die prokaryotischen Sequenzen benötigt wird.

The screenshot shows a window titled "Search results" with a search bar containing "Results of search KEGG Orthology ID: K00031". Below the search bar is a table with the following columns: Organism, KEGG ID, Gene name, UniProt ID, Selected, InterPro, and KEGG Orthology ID. The table lists various eukaryotic sequences, including Arabidopsis lyrata, Homo sapiens, and Leishmania major. The "Selected" column has checkboxes, and the "InterPro" column contains protein identifiers like IPR004790 and IPR019818. The "KEGG Orthology ID" column consistently shows K00031 for all entries.

Organism	KEGG ID	Gene name	UniProt ID	Selected	InterPro	KEGG Orthology ID
ICDH euk				<input checked="" type="checkbox"/> [9]		
Arabidopsis lyrata	aly:ARALYDRAFT_474695	(RefSeq) hypothetical protein	D7KNX5	<input type="checkbox"/>	IPR004790 Isocitrate_DH_NADP; IPR019818 IsoCif	K00031
Arabidopsis lyrata	aly:ARALYDRAFT_475696	(RefSeq) hypothetical protein	D7KT18	<input type="checkbox"/>	IPR004790 Isocitrate_DH_NADP; IPR019818 IsoCif	K00031
Arabidopsis lyrata	aly:ARALYDRAFT_488300	(RefSeq) hypothetical protein	D7M6L0	<input type="checkbox"/>	IPR004790 Isocitrate_DH_NADP; IPR019818 IsoCif	K00031
Arabidopsis lyrata	aly:ARALYDRAFT_675993	(RefSeq) hypothetical protein	D7KT19	<input type="checkbox"/>	IPR004790 Isocitrate_DH_NADP; IPR024084 IsoPro	K00031
Arabidopsis thaliana	ath:AT1G54340	(RefSeq) isocitrate dehydrogenase	Q8SLK0	<input checked="" type="checkbox"/>	IPR004790 Isocitrate_DH_NADP; IPR019818 IsoCif	K00031
Arabidopsis thaliana	ath:AT1G54930	(RefSeq) NADP-dependent isocitrate dehyd.	Q8SRZ6	<input type="checkbox"/>	IPR004790 Isocitrate_DH_NADP; IPR019818 IsoCif	K00031
Arabidopsis thaliana	ath:AT5G14590	(RefSeq) isocitrate dehydrogenase	Q8LPJ5	<input type="checkbox"/>	IPR004790 Isocitrate_DH_NADP; IPR019818 IsoCif	K00031
Homo sapiens	hsa:3417	(RefSeq) isocitrate dehydrogenase 1 (NADP)	O75874	<input checked="" type="checkbox"/>	IPR004790 Isocitrate_DH_NADP; IPR019818 IsoCif	K00031
Homo sapiens	hsa:3418	(RefSeq) isocitrate dehydrogenase 2 (NADP)	P48735	<input type="checkbox"/>	IPR004790 Isocitrate_DH_NADP; IPR019818 IsoCif	K00031
Leishmania braziliensis	lbz:LBRM_10_0310	(RefSeq) putative isocitrate dehydrogenase	A4H612	<input type="checkbox"/>	IPR004790 Isocitrate_DH_NADP; IPR019818 IsoCif	K00031
Leishmania braziliensis	lbz:LBRM_33_2820	(RefSeq) putative isocitrate dehydrogenase	A4HLW5	<input type="checkbox"/>	IPR001804 Isocitrate/isopropylmalate_DH; IPR0044	K00031
Leishmania infantum	lif:LINJ_10_0310	(RefSeq) putative isocitrate dehydrogenase	A4HUD9	<input type="checkbox"/>	IPR004790 Isocitrate_DH_NADP; IPR019818 IsoCif	K00031
Leishmania infantum	lif:LINJ_33_2680	(RefSeq) putative isocitrate dehydrogenase	A4I994	<input type="checkbox"/>	IPR001804 Isocitrate/isopropylmalate_DH; IPR0044	K00031
Leishmania major	lma:LMJF_10_0290	(RefSeq) putative isocitrate dehydrogenase	Q4QH7	<input checked="" type="checkbox"/>	IPR004790 Isocitrate_DH_NADP; IPR019818 IsoCif	K00031
Leishmania major	lma:LMJF_33_2550	(RefSeq) putative isocitrate dehydrogenase	Q4Q3T2	<input checked="" type="checkbox"/>	IPR001804 Isocitrate/isopropylmalate_DH; IPR0044	K00031
Plasmodium chabaudi	pcb:PC000375.04.0	(RefSeq) hypothetical protein	Q4XPY6	<input type="checkbox"/>	IPR004790 Isocitrate_DH_NADP; IPR024084 IsoPro	K00031
Plasmodium chabaudi	pcb:PC000412.00.0	(RefSeq) isocitrate dehydrogenase (NADP)	Q4YTE4	<input type="checkbox"/>	IPR004790 Isocitrate_DH_NADP; IPR019818 IsoCif	K00031
Plasmodium chabaudi	pcb:PC002184.00.0	(RefSeq) hypothetical protein	Q4X8E0	<input type="checkbox"/>	IPR004790 Isocitrate_DH_NADP; IPR024084 IsoPro	K00031
Plasmodium falciparum	pfa:PF13_0242	(RefSeq) isocitrate dehydrogenase (NADP)	Q8I6T2	<input checked="" type="checkbox"/>	IPR004790 Isocitrate_DH_NADP; IPR019818 IsoCif	K00031
Paramecium tetraurelia	ptm:GSPATT00005218001	(RefSeq) hypothetical protein	A0BPK2	<input checked="" type="checkbox"/>	IPR004790 Isocitrate_DH_NADP; IPR019818 IsoCif	K00031
Paramecium tetraurelia	ptm:GSPATT00009194001	(RefSeq) hypothetical protein	A0BPK2	<input type="checkbox"/>	IPR004790 Isocitrate_DH_NADP; IPR019818 IsoCif	K00031
Paramecium tetraurelia	ptm:GSPATT00020509001	(RefSeq) hypothetical protein	A0DUY0	<input type="checkbox"/>	IPR004790 Isocitrate_DH_NADP; IPR019818 IsoCif	K00031
Paramecium tetraurelia	ptm:GSPATT00022820001	(RefSeq) hypothetical protein	A0E2Y1	<input type="checkbox"/>	IPR004790 Isocitrate_DH_NADP; IPR019818 IsoCif	K00031

Abbildung 2.10: Ausgabe von HGT Downloader für die Suche nach Isocitrat-Dehydrogenasen. Das geöffnete Fenster zeigt alle eukaryotischen Sequenzen, die mit Hilfe der KEGG-Orthologie K00031 gefunden wurden.

Der Algorithmus des HGT Calculators basiert auf vier Teilen: (i) Unterschieden in den Proteindomänen, (ii) dem Verhältnis von eukaryotischen und prokaryotischen BLAST-Ergebnissen, (iii) Abweichungen vom Organismus-spezifischen GC-Gehalt und ein niedrigem Codon Adaptation Index (CAI), und (iv) eine untypische phylogenetische Zuordnung. Die phylogenetische Analyse ist hierbei auf eine geringe Auflösung beschränkt. Zu jedem Punkt werden fünf Kriterien überprüft. Wenn für eine Sequenz ein Kriterium erfüllt ist, steigt die Wertung als HGT-Ereignis. Die einzelnen Kriterien der verschiedenen Teile der Analyse sind in Tabelle 2.5 zusammengefasst.

Der Schwellwert für die HGT-Wertung wurde von Gensequenzen hergeleitet, die keine HGT-Ereignisse sind und mit einem Wert von 45 definiert. Einige Kriterien müssen erfüllt sein, um ein HGT-Ereignis zu belegen.

1. Die Sequenz sollte zu einer anderen Proteinuntergruppe gehören als die übrigen eukaryotischen Sequenzen des Datensatzes.
2. Die Proteindomänen dieser Sequenz sollten häufiger in Prokaryoten vorkommen als in Eukaryoten.
3. Über 50 % der BLAST-Ergebnisse sollten aus Prokaryoten stammen.
4. Der Alien-Index (AI) für diese Sequenz sollte größer als 45 sein.
5. Die Sequenz sollte sich auf allen drei Ebenen der Neighbor-Joining-Analyse (einzeln / Level1-Baum, kleiner / Level2-Baum, großer / Level3-Baum; Abbildung 2.14) in nächster Nachbarschaft zu einer prokaryotischen Sequenz befinden.
6. Der Level3-Baum sollte in Teilbäume unterteilt sein, die zum einen eukaryotische und prokaryotische Sequenzen und zum anderen unterschiedliche Proteinunterarten

voneinander trennen. Auf diese Weise werden sehr alte und bei Prokaryoten und Eukaryoten konservierte Gene ausgeschlossen.

Die Summe all dieser Kriterien ist 45 für eine Standardgewichtung in der Bewertung. Ein Testdatensatz aus zehn ursprünglichen Genen belegt, dass Gene, die nicht das Resultat von horizontalem Gentransfer sind, sich in einer Proteinunterart gruppieren. Alle HGT-Werte sind null und die Zwischenwerte sind kleiner als 45 (Abbildung 2.6). Generell werden ursprüngliche Gene mit dem Mangel an prokaryotischen BLAST-Treffern und der Struktur der Neighbor-Joining-Bäume als mögliches HGT-Ereignis ausgeschlossen. Diese Ergebnisse werden hinter dem HGT-Wert mit einem 'b' beziehungsweise einem 't' vermerkt.

	ACO_1	AHCY_1	ENO_1	GUK_1	MAT_1	NDK_1	RRS_1	SHMT_1	TPI_1	TYMS_1
<i>Tetrahymena</i>	0 (30) b t	0 (30) b t	0 (30) b t	0 (15) b t	0 (25) b t	0 (20) b t	0 (15) b t	0 (10) b t	0 (30) b t	0 (10) b t
<i>Paramecium</i>	0 (25) b t	0 (25) b t	0 (20) b t	0 (10) b t	0 (30) b t	0 (35) b t	0 (15) b t	0 (20) b t	0 (20) b t	0 (10) b t
<i>Strigomonas</i>	0 (25) b t	0 (30) b t	0 (30) b t	0 (25) b t	0 (25) b t	0 (25) b t	0 (20) b t	0 (25) b t	0 (25) b t	0 (25) b t
<i>Leishmania</i>	0 (20) b t	0 (20) b t	0 (25) b t	0 (20) b t	0 (20) b t	0 (20) b t	0 (20) b t	0 (15) b t	0 (20) b t	0 (20) b t
<i>Trypanosoma</i>	0 (30) b t	0 (15) b t	0 (20) b t	0 (10) b t	0 (10) b t	0 (20) b t	0 (20) b t	0 (10) b t	0 (10) b t	0 (10) b t
<i>Toxoplasma</i>	0 (20) b t	0 (15) b t	0 (25) b t	0 (10) b t	0 (20) b t	0 (25) b t	0 (20) b t	0 (20) b t	0 (20) b t	0 (20) b t
<i>Plasmodium</i>	0 (20) b t	0 (30) b t	0 (25) b t	0 (10) b t	0 (25) b t	0 (25) b t	0 (20) b t	0 (30) b t	0 (25) b t	0 (15) b t
<i>Saccharomyces</i>		0 (10) b t	0 (20) b t	0 (15) b t	0 (25) b t	0 (20) b t	0 (15) b t	0 (10) b t	0 (20) b t	0 (15) b t
<i>Arabidopsis</i>	0 (10) b t	0 (10) b t	0 (15) b t			0 (15) b t	0 (20) b t	0 (15) b t	0 (15) b t	0 (15) b t
<i>Homo</i>	0 (25) b t	0 (20) b t	0 (15) b t	0 (25) b t	0 (15) b t	0 (10) b t	0 (20) b t	0 (25) b t	0 (20) b t	0 (10) b t

Tabelle 2.6: HGT-Werte für Kontrollgene, die keine HGT-Ereignisse sind.

Darstellung in der Tabelle: HGT-Wert (HGT-Zwischenwert)

b: Die BLAST-Ergebnisse sind nicht überzeugend.

t: Die Topologie der Neighbor-Joining-Bäume liefert keine Belege für horizontalen Gentransfer.

(Tabelle nach Ellenberger, Siegmund et al. (2016))

Im ersten Teil der Analyse werden für Eukaryoten untypische Sequenzen auf Proteinebene untersucht. Abweichungen in der Sequenzlänge und Unterschiede in den Proteindomänen werden ermittelt. Ein deutlicher Unterschied in der Sequenzlänge ist ein Hinweis auf horizontalen Gentransfer. Wenn die Sequenz länger ist, als andere eukaryotische Sequenzen für dieses Gen, wurden vielleicht Teile einer fremden Sequenz eingebaut. Ist die Sequenz dagegen kürzer als erwartet, wurde das Gen möglicherweise mit einer prokaryotischen Version ausgetauscht. Dieser fehlen die eukaryotischen Proteindomänen und regulatorischen Schnittstellen, die es länger machen würden. Für alle auftretenden Proteindomänen wird überprüft, wie häufig sie in Prokaryoten und Eukaryoten vorkommen. Zu diesem Zweck werden die zugehörigen InterPro-Identifikationsnummern (Jones et al., 2014) (www.ebi.ac.uk/interpro) benutzt. Wenn diese Proteindomäne nur bei wenigen Eukaryoten zu finden ist, ist das ein weiterer Hinweis auf horizontalen Gentransfer.

Der zweite Teil der Analyse wertet das Eukaryoten/Prokaryoten-Verhältnis von BLAST-Ergebnissen aus. BLAST (Altschul et al., 1990) wird über HGT Downloader in UniProt

gestartet. Die Zuordnung der 250 besten BLAST-Ergebnisse in die drei Gruppen Archaeen, Bakterien und Eukarya zeigt horizontalen Gentransfer an. Eine große Anzahl an bakteriellen Sequenzen in den BLAST-Ergebnissen einer eukaryotischen Sequenz wird als Hinweis auf horizontalen Gentransfer von Bakterien zu Eukaryoten interpretiert.

Organism	UniProt ID	Status
Strigomonas	S9UPN9	204
Leishmania	Q4Q3T2	Finished [200.0]
Leishmania	Q4QHI7	195
Arabidopsis	Q9SLK0	Finished [0.0]
Homo	O75874	0
Plasmodium	Q8I6T2	0
Paramecium	A0BPK2	91
Saccharomyces	P21954	Finished [0.0]
Tetrahymena	Q23K14	0
Toxoplasma	Q5UCG0	172
Strigomonas	S9UPW7	0
Trypanosoma	A0A061IWC2	108

→ Ein Alien-Index von 200 belegt HGT.

→ Ein Alien-Index von 0 widerlegt HGT.

Abbildung 2.11: Ausgabe von HGT Downloader für die BLAST-Suche in UniProt und Berechnung des Alien-Index für Isocitrat-Dehydrogenasen.

Der Alien-Index für die ICDH_1-Variante der Isocitrat-Dehydrogenase aus *Leishmania major* (UniProt: Q4Q3T2) ist 200 und gibt ein deutliches HGT-Signal.

Gladyshev et al. (2008) präsentierten eine Methode zur Auswertung von BLAST-Ergebnissen zum Auffinden von horizontalem Gentransfer in bdelloiden Rädertieren. Der dort vorgestellte Alien-Index (AI) wurde an den spezifischen Transfer von Prokaryoten zu Protozoen angepasst:

$$AI = \log[(\text{Bester E-Wert für Eukaryoten}) + E^{-200}] - \log[(\text{Bester E-Wert für Prokaryoten}) + E^{-200}] \quad (2.13)$$

Der Alien-Index wird nach Abschluss der BLAST-Suche automatisch berechnet (Abbildung 2.11) und in der ersten Zeile einer Textdatei gespeichert, die alle BLAST-Treffer der Ausgangssequenz beinhaltet. Diese Kopfzeile speichert zusätzlich die jeweilige Anzahl von Treffern in den drei Gruppen Archaeen, Bakterien und Eukarya, ob der Organismus, aus dem die Ausgangssequenz stammt, der einzige eukaryotische Organismus ist, ob der E-Wert und die Sequenzabdeckung aussagekräftig sind, sowie die mögliche Quelle der fremden DNA. Ein Alien-Index, der größer gleich 45 ist, wird als Hinweis auf horizontalen Gentransfer interpretiert.

HGT Downloader erzeugt eine Statistik der BLAST-Ergebnisse. In einem Kreuzdiagramm wird die Anzahl der prokaryotischen BLAST-Treffer festgehalten. In einem Blockdiagramm werden die Mittelwerte und Standardabweichungen eingetragen. Sofern es für einen Organismus mehrere Sequenzen gibt, zeigen sich die Unterschiede zu HGT-Sequenzen (Abbildung 2.12). Die Diagramme lassen sich als Abbildung speichern.

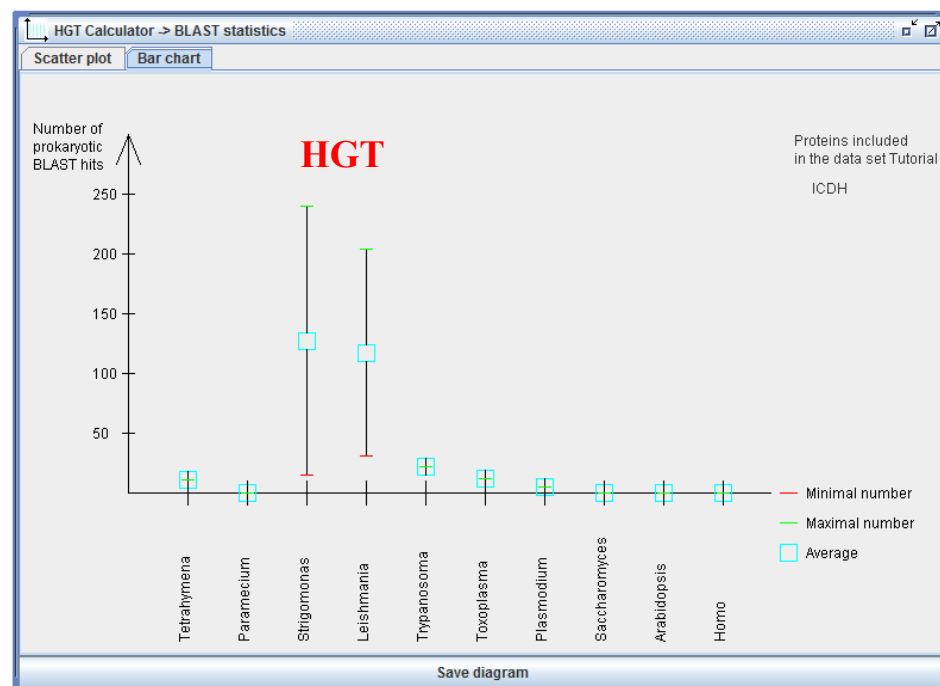
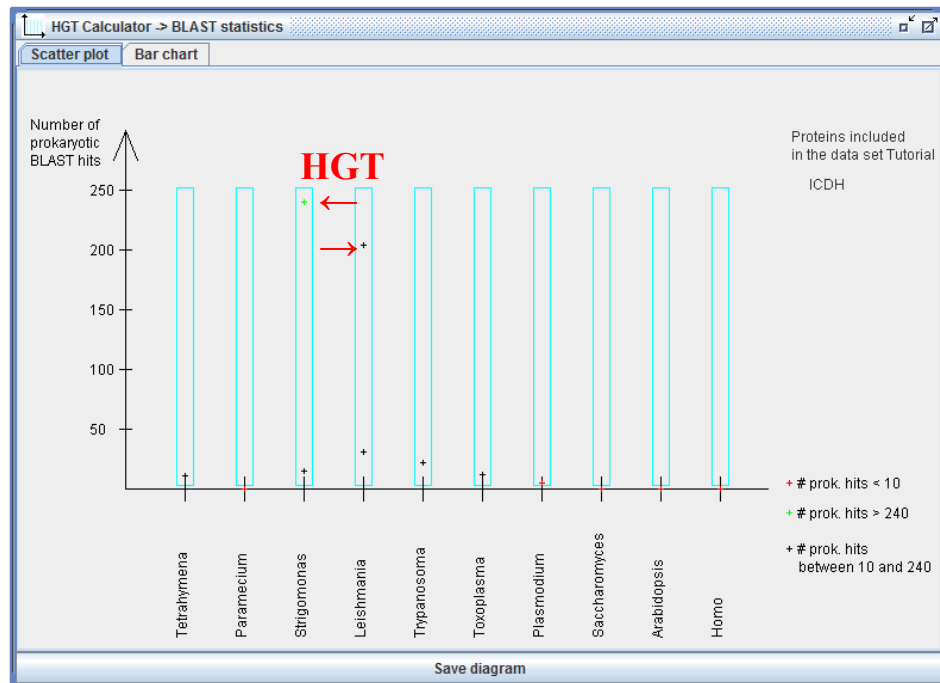


Abbildung 2.12: Ausgabe von HGT Downloader für die BLAST-Statistik für Isocitrat-Dehydrogenasen. Die ICDH₁-Varianten der Isocitrat-Dehydrogenasen aus *Leishmania major* und *Strigomonas culicis* geben ein deutliches HGT-Signal.

BLAST-Suchen sind allerdings nur erste Indikatoren auf horizontalen Gentransfer. Sie beschreiben phylogenetische Beziehungen nur ungenau (Koski und Golding, 2001) und sollten deshalb nicht als einzige Methode zur Identifizierung von HGT-Ereignisse benutzt werden. Im HGT Calculator ist die Auswertung der BLAST-Ergebnisse nur ein Baustein, der mit anderen Analyseverfahren kombiniert wird.

Der dritte Teil der Analyse untersucht den GC-Gehalt der Sequenzen. In vielen Organismen weicht der GC-Gehalt von den theoretischen 50 % ab und ist besonders für Bakterien ein Organismen-spezifisches Merkmal. Der typische GC-Gehalt der untersuchten Organismen stammt aus der Codon-Usage-Datenbank (Nakamura et al., 2000). Der GC-Gehalt innerhalb einer Spezies variiert zum Teil stark. Die codierenden Bereiche sind allerdings in ihrem GC-Gehalt und der Häufigkeit, mit der bestimmte Codons auftreten, vergleichbar. Sequenzen, die über horizontalen Gentransfer in einen Organismus eingewandert sind, behalten ihre ursprünglichen Sequenzeigenschaften eine gewisse Zeit lang bei, bis sie sich an den neuen Organismus und die dort vorkommenden Verhältnisse anpassen. Dadurch sind sie von den eigenen Genen unterscheidbar (Baran und Ko, 2008). Dieses Signal ist natürlich für rezente Ereignisse am stärksten. Doch obwohl diese Detektionsmethode sehr hilfreich sein kann, wurde gezeigt, dass die Ergebnisse, die auf diese Art gewonnen werden, den Ergebnissen aus anderen Analysen zum Teil widersprechen (Ragan, 2001). Deshalb ist dieses Verfahren nur ein Baustein des HGT Calculators und wird von anderen Techniken unterstützt und gegebenenfalls korrigiert.

Der Codon Adaptation Index (CAI) (Sharp und Li, 1987) ist ein Parameter, der die relative Anpassungsfähigkeit der Verwendung von Codons eines vorgegebenen Gens im Vergleich zur Verwendung von Codons in stark exprimierten Genen beschreibt. Die relative Anpassungsfähigkeit jedes Codons ist das Verhältnis der Verwendung jedes Codons zu der des am häufigsten auftretenden Codons für dieselbe Aminosäure. Der CAI ist definiert als der geometrische Mittelwert dieser Werte, der die relative Anpassungsfähigkeit beschreibt. Nicht synonyme Codons und Stopcodons, in Abhängigkeit vom jeweiligen genetischen Code eines Organismus, werden von der Berechnung ausgeschlossen. CAI-Werte liegen zwischen null und eins, wobei höhere Werte für einen größeren Anteil des häufigsten Codons für eine Aminosäure stehen. Niedrige CAI-Werte werden als Hinweis auf horizontalen Gentransfer interpretiert. Die CAI-Werte der Sequenzen werden im HGT Calculator nach der Implementierung von Xia (2007) berechnet. Tabellen über die typische Verwendung von Codons innerhalb eines Organismus werden von der Codon-Usage-Datenbank (Nakamura et al., 2000) heruntergeladen.

Der vierte Teil der Analyse verwendet einfache Baum-basierte Methoden um HGT-Ereignisse zu erkennen. Der eindeutigste Beweis für ein HGT-Ereignis liegt in dem Auftreten von unerwarteten Beziehungen zwischen einer oder mehrerer Spezies innerhalb einer gewurzelten Baum-Topologie, die aus gut ausgewählten Sequenzen aufgebaut ist und anerkannte Spezies-Verwandtschaften darstellt (Andersson, 2005). HGT-Ereignisse

finden sich außerhalb dieser traditionellen Baum-Strukturen. Deshalb reduziert HGT Calculator die phylogenetische Analyse auf die Verwendung als interne Datenstruktur. Dabei werden nicht-gewurzelte Bäume verwendet, denn horizontaler Gentransfer verwandelt Bäume in Netzwerke. Eine eukaryotische Sequenz, die zwischen prokaryotischen Sequenzen eingebettet liegt, erkennt man auch in Netzwerk-Topologien. Rezente HGT-Ereignisse führen erwartungsgemäß zu anderen Baum-Strukturen, wenn die taxonomische Abdeckung ausreichend ist und Transferereignisse relativ häufig auftreten. Ältere endosymbiotische Gentransferereignisse sind erwartungsgemäß als einzelner eukaryotischer Teilbaum zu erkennen, der sich möglichst innerhalb von Sequenzen aus Cyanobakterien oder α -Proteobakterien befindet. Echte HGT-Ereignisse sollten sich als polyphyletische Eukaryoten darstellen und sich möglichst innerhalb eines prokaryotischen Teilbaums befinden (Andersson, 2005). Da HGT Calculator keine detaillierte Phylogenie der untersuchten Organismen erstellen soll, beschränkt sich die Analyse in diesem Teil des Programms auf einfache ungewurzelte Neighbor-Joining-Bäume. Diese sind ausreichend, um einzelne Gruppen zu identifizieren, die ausschließlich aus eukaryotischen oder prokaryotischen Sequenzen bestehen. Horizontaler Gentransfer zeigt sich hier in einzelnen eukaryotischen Sequenzen, die sich innerhalb des prokaryotischen Teilbaums befinden.

Der neue Ansatz zum Auffinden von HGT-Ereignissen in ungewurzelten phylogenetischen Bäumen, der in HGT Calculator erstmals angewendet wird, arbeitet als Multi-Level-Analyse auf mehreren Bäumen unterschiedlicher Auflösung, das heißt, mit einer unterschiedlichen Anzahl an Sequenzen.

Der erste Baum wird als *einzelner* Baum oder Level1-Baum bezeichnet. Er besteht aus den Sequenzen eines Protozoons und zusätzlichen Sequenzen multi-zellulärer Organismen als Randgruppe. Diese Sequenzen stammen meistens aus einem Pilz (*Saccharomyces cerevisiae*), einer Pflanze (*Arabidopsis thaliana*), und einem Metazoon (*Homo sapiens*). Sollte ein Gen nicht in Pflanzen vorkommen, muss eine Sequenz aus einem weiteren Pilz ausgewählt werden. Entsprechend muss eine alternative Sequenz gefunden werden, wenn das gesuchte Gen nicht in Pilzen oder Metazoa vorkommt. Hinzu kommt jeweils eine bakterielle Sequenz für jede eukaryotische Proteinunterart, die von den Proteindomänen der Sequenzen definiert wurden. So ein einfacher Baum wird für jedes Protozoon einzeln erzeugt. Ein weiterer Baum beinhaltet nur die Randgruppe und die bakteriellen Sequenzen zur Kontrolle, um EGT-Ereignisse auszuschließen. Alle Level1-Bäume lassen sich direkt in HGT Calculator darstellen (Abbildung 2.13). Dabei werden Zweige, die ein HGT-Signal liefern farbig markiert. Die Bäume lassen sich als Abbildung speichern. Die Ansicht lässt sich zwischen proportionalen und nicht proportionalen Astlängen wechseln. Für eine bessere Abbildung lässt sich die Größe der Bäume anpassen.

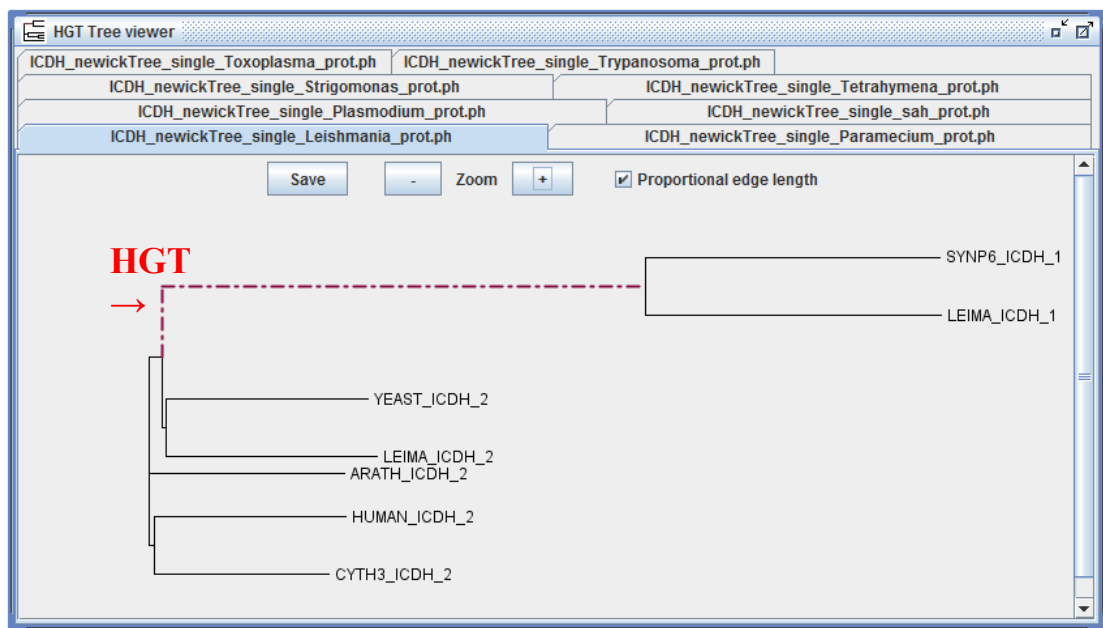


Abbildung 2.13: Ausgabe von HGT Downloader für die Level1-Bäume der Isocitrat-Dehydrogenasen. Die ICDH₁-Variante der Isocitrat-Dehydrogenase aus *Leishmania major* erzeugt ein deutliches HGT-Signal.

Der zweite Baum wird als *kleiner* Baum oder Level2-Baum bezeichnet. Er besteht aus allen eukaryotische Sequenzen und einer bakteriellen Sequenz der unterschiedlichen Proteinunterarten.

Der dritte Baum wird als *großer* Baum oder Level3-Baum bezeichnet. Er enthält alle Sequenzen des Datensatzes für ein bestimmtes Gen.

Die Neighbor-Joining-Bäume werden mit ClustalW (Larkin et al., 2007) erzeugt, das von HGT Calculator gestartet wird. Abbildung 2.14 zeigt die Bäume der verschiedenen Level am Beispiel der Isocitrat-Dehydrogenase.

Die Bäume werden auf die Position der eukaryotischen Sequenzen hin untersucht. Als *positive direkte Nachbarn* werden eukaryotische Sequenzen bezeichnet, die sich in nächster Nachbarschaft zu einer prokaryotischen Sequenz befindet und deren Vaterknoten eine definierte Distanz (0,07) zum eukaryotischen Teilbaum einhält. Dieser Schwellwert wurde aus Testdatensätzen gelernt. Die Korrektheit der Baumtopologie wird im Level3-Baum überprüft, indem die Teilbäume untersucht werden. Lässt sich der Baum in eukaryotische und prokaryotische Teilbäume zerlegen und existiert auch für jede Proteinunterart ein eigener Teilbaum, ist die Baumtopologie akzeptabel und der HGT-Wert erhöht sich. Um als HGT-Ereignis gewertet zu werden, muss sich die Sequenz direkt neben einer bakteriellen Sequenz im Baum befinden.

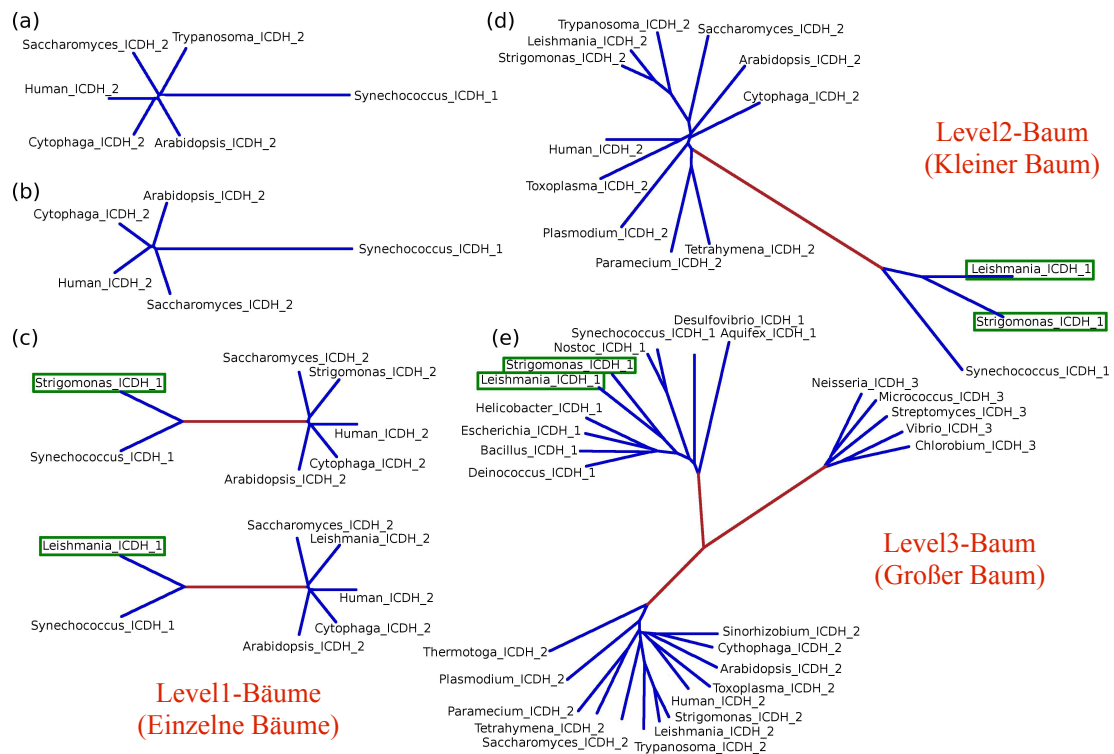


Abbildung 2.14: Neuer Ansatz für die Detektion von HGT-Ereignissen in ungewurzelten Neighbor-Joining-Bäumen mit Multi-Level-Analyse.

(a) Level1-Baum für *Trypanosoma rangeli* ohne HGT-Signal

(b) Level1-Baum für multi-zelluläre Referenzorganismen ohne HGT-Signal

(c) Level1-Bäume für *Strigomonas culicis* und *Leishmania major* mit HGT-Signal

(d) Level2-Baum und

(e) Level3-Baum mit HGT-Signal für ICDH_1-Variante der Isocitrat-Dehydrogenasen in *Leishmania major* und *Strigomonas culicis* (UniProt ID: Q4Q3T2, S9UPN9, gekennzeichnet mit grünen Kästchen).

Die roten Linien verdeutlichen die für HGT-Ereignisse erwartete Baum-Topologie:

Zwei getrennte Teilbäume für eukaryotische und prokaryotische Sequenzen, mit einzelnen eukaryotischen Sequenzen, die sich innerhalb des prokaryotischen Teilbaums befinden.

(Abbildung nach Ellenberger, Sigmund et al. (2016))

Die Ergebnisse der einzelnen Analysen werden am Schluss in einem HGT-Wert zusammengefügt. Einmalige HGT-Ereignisse, die ihren Ursprung in einem Endosymbionten haben, wie zum Beispiel Homoserin-Dehydrogenase aus *Strigomonas culicis* (UniProt: S9UCM8), ergeben einen HGT-Wert, der deutlich über dem Schwellwert von 45 liegt. Sie treten als einziger HGT-Wert auf, der für dieses Protein über null liegt. HGT-Ereignisse, die innerhalb eines Vorfahren noch vor der Differenzierung in verschiedene taxonomische Linien stattgefunden haben, ergeben mehrere HGT-Werte über 30. Beispiele hierfür sind Fumarat-Hydratase oder Isopentenylidiphosphat-Isomerase, die sowohl in *Leishmania*- als auch in *Trypanosoma*-Stämmen vorkommen. Lange zurückliegende, mehrfach auftretende HGT-Ereignisse oder EGT-Ereignisse ergeben HGT-Zwischenwerte kleiner als 30, die zwar von der Baumtopologie als mögliches HGT-Ereignis zurückgewiesen,

allerdings von den BLAST-Ergebnissen unterstützt werden. Beispiele hierfür sind Fer-rochelata-se oder Glutamat-Dehydrogenase, die sowohl in Trypanosomatiden (Opperd-oes und Michels, 2007) als auch in Apicomplexa (Huang, Mullapudi, Sicheritz-Ponten et al., 2004) als HGT-Ereignisse beschrieben sind.

Die HGT-Werte werden in einem extra Fenster, dem ResultManager, angezeigt. Abbildung 2.15 zeigt die HGT-Werte für Isocitrat-Dehydrogenasen. Potentielle HGT-Ereignisse lassen sich mit Hilfe des ResultMangers bereits anhand einer Farbcodierung auf den ersten Blick identifizieren. Grüne Einträge sind einzeln auftretende rezente HGT-Ereignisse mit einem HGT-Wert über 45. Blaue Einträge sind ältere Ereignisse, die vor der Spezialisierung in unterschiedliche taxonomische Linien in einem gemeinsamen Vorfahren stattgefunden haben. Oder es handelt sich um Gene, die von mehreren voneinander unabhängigen HGT-Ereignissen in unterschiedlichen miteinander nicht nah verwandten Organismen entstanden sind. Alle anderen Farben können vom Nutzer selbst definieren werden. Die entsprechenden Einträge werden über das Auswahlmenü markiert. Der ResultManager markiert den ausgewählten Eintrag in allen Tabellen. So erhält der Nutzer einen Überblick über die Eigenschaften dieser Sequenz: UniProt-Identifikationsnummer, Sequenzlänge, Gen-Identifikationsnummer, Proteindomänen, BLAST-Ergebnisse, GC-Gehalt und Codon Adaptation Index, sowie eine binär codierte Kriterienliste (1: Kriterium ist erfüllt, 0: Kriterium ist nicht erfüllt).

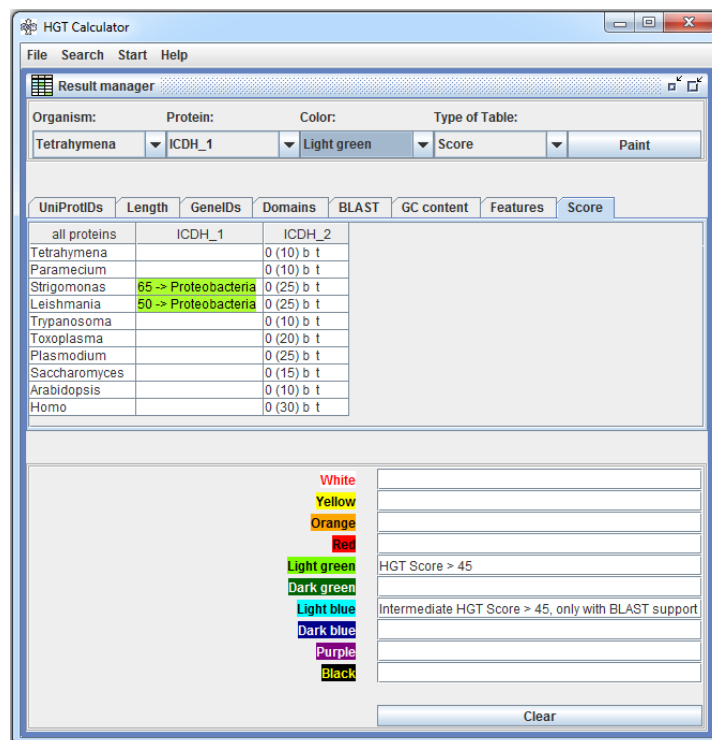


Abbildung 2.15: Ausgabe von HGT Calculator für Isocitrat-Dehydrogenasen im ResultManager. Die ICDH_1-Varianten der Isocitrat-Dehydrogenasen aus *Leishmania major* und *Strigomonas culicis* erzeugen ein deutliches HGT-Signal. Die HGT-Werte sind größer als 45.

2.9.2 Genvorhersage für *Absidia glauca* und *Parasitella parasitica*:**ZygoViewer**

ZygoViewer ist eine graphische Oberfläche (Abbildung 2.16), die entwickelt wurde, um bereits vor der Veröffentlichung der für die beiden sequenzierten Genome von *Absidia glauca* und *Parasitella parasitica* vorhergesagten Gene in ENA (European Nucleotide Archive) die Arbeit mit diesen Daten zu ermöglichen. Mit Hilfe des ZygoViewers lässt sich gezielt nach Genen innerhalb dieser beiden Organismen suchen. Zur Auswahl stehen dafür verschiedene Schlüsselwörter wie Genname, Genfunktion, Proteindomäne oder Lage auf einem Scaffold. Eine Tabelle mit allen gefundenen oder nur einigen ausgewählten Genen lässt sich für weitere Aufgaben als Textdatei speichern. Sequenzen lassen sich mit einem Klick mit der rechten Maustaste direkt anzeigen. Zu sehen sind neben Nukleotid- und Proteinsequenz auch die einzelnen Exons des ausgewählten Gens. Eine weitere Option ist die Ansicht in Artemis (Rutherford et al., 2000), die eine Betrachtung des genomischen Kontextes ermöglicht.

Eines der Programmfenster dient der Ansicht der TSP1-Modelle für *Absidia glauca* und *Parasitella parasitica* (Abbildung 2.17). Über eine Auswahlleiste lassen sich die verschiedenen TSP1-Proteine aus *Absidia glauca* und *Parasitella parasitica* als Rezeptor oder Ligand auswählen. Die Dimere werden in unterschiedlichen Darstellungsvarianten und Blickwinkeln aufgelistet. So zum Beispiel als Oberflächenmodell mit Blick auf die Trisporoid-Bindestelle oder als Bandmodell mit eingefärbten Strukturelementen mit Blick auf die Dimerisierungszone. Nach der Auswahl eines Dimers werden die entsprechenden Einträge in den Tabellen markiert, die die Eigenschaften der Dimere darstellen. Anhand von Cluster-Anzahl, Cluster-Größe, ClusPro-Wert und möglicher Enzymaktivität lassen sich die einzelnen Dimere miteinander vergleichen und auswerten.

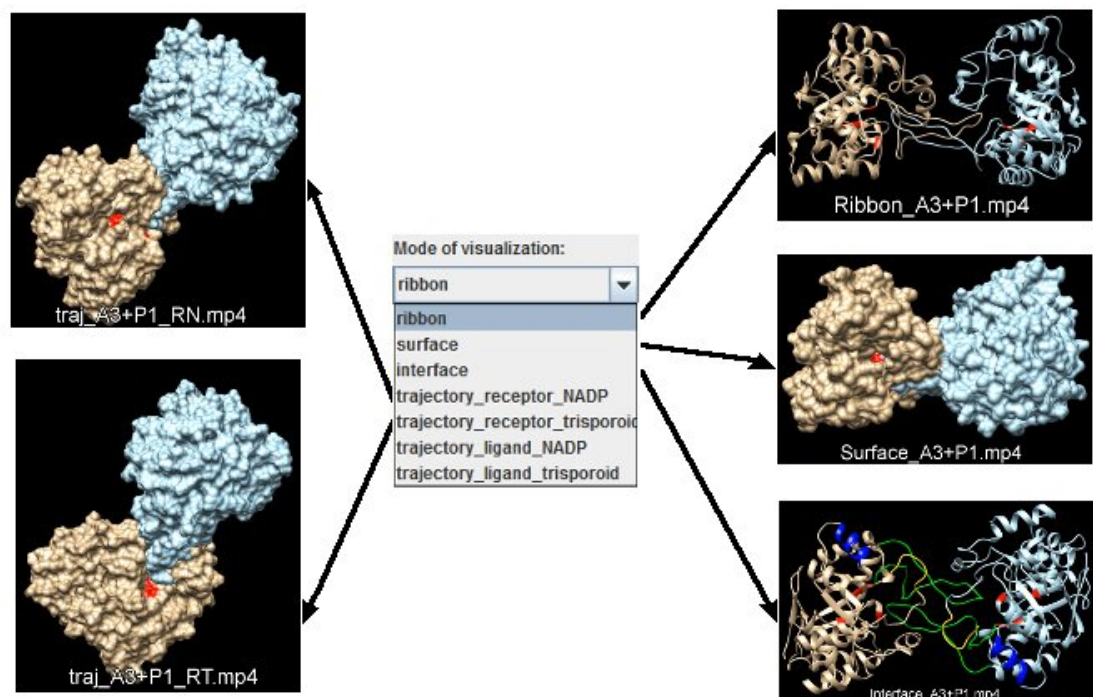
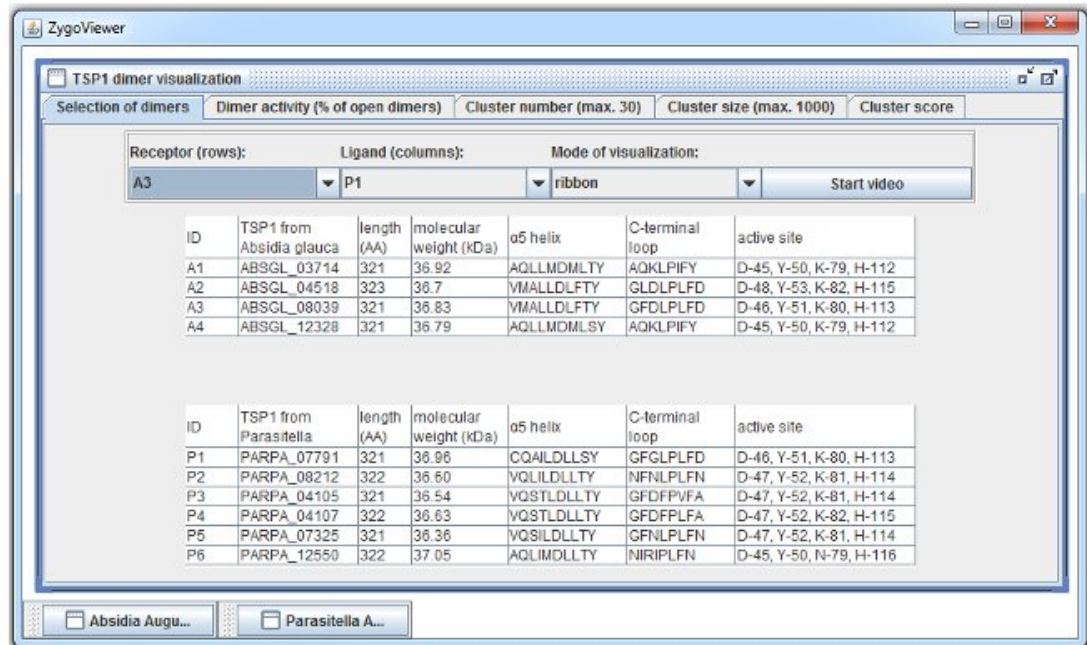


Abbildung 2.17: ZygoViewer: Visualisierung der TSP1-Dimer-Modelle für *Absidia glauca* und *Parasitella parasitica*. Die modellierten Dimere werden in kurzen Videos in verschiedenen Darstellungen und aus unterschiedlichen Blickwinkeln präsentiert.

3 Ergebnisse

3.1 Vergleich der Genome von Wirt und Parasit

Der Pilz *Parasitella parasitica* ist ein Fusionsparasit. An der Kontaktstelle zwischen Wirt und Parasit kommt es bei der Ausbildung von Infektionsstrukturen zu einer Übertragung von Zellmaterial aus dem Parasiten hinein in die Wirtshyphe. Die Integration einzelner Gene aus dem Parasiten in das Genom des Wirtes wurde für *Absidia glauca* nachgewiesen (Burmester et al., 2013). Um den Anteil an fremdem genetischen Material im Wirtsgenom zu ermitteln, wurden die Genome von Wirt (*Absidia glauca* CBS 101.48 RVII-324 met⁻) und Parasit (*Parasitella parasitica* CBS 412.66 NGI-315 ade⁻) sequenziert und miteinander verglichen.

3.1.1 Das Genom von *Absidia glauca*

Das Genom von *Absidia glauca* (ENA Projektnummer: PRJEB13349) hat eine Größe von rund 49 Mbp und einen GC-Gehalt von 45 %. In Introns liegt der GC-Gehalt nur bei 29 %. Innerhalb von codierenden Bereichen liegt der GC-Gehalt höher, bei rund 48 %. Insgesamt wurden 15 594 Gene vorhergesagt. Einige dieser Gene sind kurze Fragmente mit einer Länge von nur rund 40 bp. Das längste Gen hat eine Länge von 20 640 bp. Die durchschnittliche Genlänge beträgt 1 275 bp. Ein Gen hat im Durchschnitt etwa 4,5 Introns. Das Genom enthält 232 tRNAs, 13 von ihnen beinhalten ein Intron. Das Genom enthält fast 63 000 repetitive Sequenzen aus 16 Familien transponierbarer Elemente.

Die mitochondriale DNA von *Absidia glauca* hat eine Länge von circa 63 kb und einen GC-Gehalt von 28 %. Sie enthält die für Pilze typischen mitochondrialen Gene (Aguileta et al., 2014), drei ATP-Synthase-Untereinheiten (*atp6*, *atp8*, *atp9*), Apocytochrom-b (*cob*), drei Cytochrom-c-Oxidase-Untereinheiten (*cox1*, *cox2*, *cox3*) und sieben NADH-Dehydrogenase-Untereinheiten (*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*). Hinzu kommen die kleine und die große rRNA-Untereinheit (*rns*, *rnl*), eine Ribonuklease-P-RNA (*rnpB*) und 24 tRNAs. Die mtDNA von *Absidia glauca* besitzt acht GIY-YIG- und vier LAGLIDADG-Endonukleasen. Fünf der acht GIY-YIG- und alle der vier LAGLIDADG-Endonukleasen befinden sich in einem Intron. Das *cox1*-Gen beinhaltet die meisten Introns. In diesem Gen liegen fünf der insgesamt 19 Introns.

Die für *Absidia glauca* vorhergesagten Gene werden im ZygoViewer (siehe Abschnitt 2.9.2) aufgelistet und wurden im European Nucleotide Archive (ENA) und auf GitHub veröffentlicht. Das mitochondriale Genom von *Absidia glauca* wurde in *Genome Announcements* (Ellenberger, Burmester und Wöstemeyer, 2016a) und in GenBank (KU196782) veröffentlicht.

3.1.2 Das Genom von *Parasitella parasitica*

Das Genom von *Parasitella parasitica* (ENA Projektnummer: PRJEB7124) hat eine Größe von rund 45 Mbp und einen GC-Gehalt von 39 %. In Introns liegt der GC-Gehalt nur bei 24 %. Innerhalb von codierenden Bereichen liegt der GC-Gehalt höher, bei rund 44 %. Insgesamt wurden 14 507 Gene vorhergesagt. Einige dieser Gene sind kurze Fragmente mit einer Länge von nur rund 40 bp. Das längste Gen hat eine Länge von 18 387 bp. Die durchschnittliche Genlänge beträgt 1 226 bp. Ein Gen hat im Durchschnitt etwa 3,6 Introns. Das Genom enthält 161 tRNAs, neun von ihnen beinhalten ein Intron. Das Genom enthält fast 49 000 repetitive Sequenzen aus 20 Familien transponierbarer Elemente.

Die mitochondriale DNA von *Parasitella parasitica* hat eine Länge von circa 83 kb und einen GC-Gehalt von 30 %. Sie enthält die für Pilze typischen mitochondrialen Gene (Aguileta et al., 2014), drei ATP-Synthase-Untereinheiten (*atp6*, *atp8*, *atp9*), Apocytochrom-b (*cob*), drei Cytochrom-c-Oxidase-Untereinheiten (*cox1*, *cox2*, *cox3*) und sieben NADH-Dehydrogenase-Untereinheiten (*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*). Hinzu kommen die kleine und die große RNA-Untereinheit (*rns*, *rnl*), eine Ribonuklease-P-RNA (*rnpB*) und 26 tRNAs. Die mtDNA von *Parasitella parasitica* enthält im Vergleich zu anderen Zygomyceten besonders viele Endonukleasen, sieben GIY-YIG- und 19 LAGLIDADG-Endonukleasen. Fünf der sieben GIY-YIG- und 18 der 19 LAGLIDADG-Endonukleasen befinden sich in einem Intron. Das *cox1*-Gen beinhaltet die meisten Introns. In diesem Gen liegen acht der insgesamt 27 Introns.

Die für *Parasitella parasitica* vorhergesagten Gene werden im ZygoViewer (siehe Abschnitt 2.9.2) aufgelistet und wurden im European Nucleotide Archive (ENA) und auf GitHub veröffentlicht. Das mitochondriale Genom von *Parasitella parasitica* wurde in *Genome Announcements* (Ellenberger et al., 2014) und in GenBank (KM382275) veröffentlicht.

3.1.3 Sequenzähnlichkeit der für *Absidia glauca* vorhergesagten Gene

Die Funktion der für *Absidia glauca* vorhergesagten Gene wurde mit BLAST in NCBI ermittelt. Abbildung 3.1 zeigt, aus welchen Organismen die gefundenen BLAST-Treffer stammen. *Absidia glauca* gehört zur Familie der Cunninghamellaceae. Unter den drei besten BLAST-Treffern finden sich allerdings viele Sequenzen aus *Mucor*-artigen Zygomyceten.

Für über 1 400 Gene stammt der erste BLAST-Treffer aus *Parasitella parasitica*. Einige dieser Sequenzen sind möglicherweise das Ergebnis von horizontalem Gentransfer zwischen Wirt und Parasit. Für 309 der von diesen Genen codierten Proteine wurden keine Proteindomänen gefunden. Viele der übrigen Gene codieren für transponierbare Elemente, wie Integrasen, Nukleasen und Transposasen, und weitere repetitive Sequenzen.

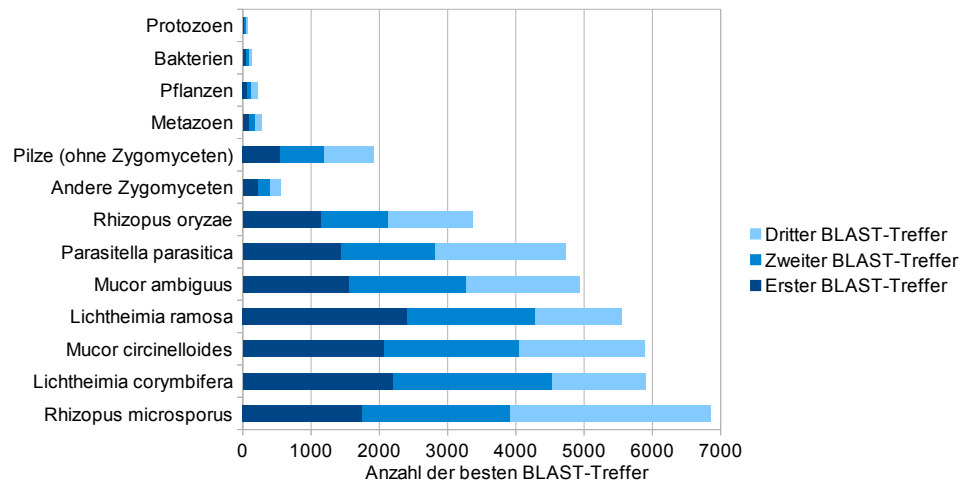


Abbildung 3.1: Ursprung der besten drei BLAST-Treffer für Gene aus *Absidia glauca*.

Obwohl *Absidia glauca* nicht zu den *Mucorales* gehört, stammen viele BLAST-Treffer aus *Mucor*-Stämmen oder *Mucor*-artigen Zygomyceten. *Absidia glauca* gehört zur Familie der Cunninghamellaceae.

Andere Gene mit einer hohen Sequenzähnlichkeit zu Genen aus *Parasitella parasitica* codieren zum Beispiel für Aktin, ribosomale Proteine, Translationsinitiationsfaktoren und Kinasen.

3.1.4 Analyse der mitochondrialen DNA von Zygomyceten

Für die Analyse mitochondrialer DNA von Zygomyceten wurden die Chondriome von *Absidia glauca* CBS 101.48 (Absgl), *Absidia repens* NRRL 1336 (Absre), *Backusella circina* FSU 941 (Bacci), *Glomus sp.* DAOM 240422 (Glosp), *Lichtheimia corymbifera* FSU 9682 (Licco), *Lichtheimia ramosa* (Licra), *Mortierella verticillata* NRRL 6337 (Morve), *Mucor circinelloides* CBS 277.49 (Mucci), *Parasitella parasitica* CBS 412.66 (Parpa), *Parasitella parasitica* NRRL 2501 (Parpar), *Phycomyces blakesleeana* NRRL 1555 (Phybl), *Rhizopus microsporus* ATCC 11559 (Rhimi), *Rhizopus oryzae* DAOM 148428 (Rhior), *Umbelopsis ramanniana* AG (Umbra) und *Zancudomyces culisetiae* 18-3 (Zancu) miteinander verglichen. Aus den Ergebnissen entstand eine Publikation in *Mitochondrial DNA Part A* (Ellenberger, Burmester und Wöstemeyer, 2016b).

Die mtDNAs von Zygomyceten enthalten drei ATP-Synthase-Untereinheiten (*atp6*, *atp8*, *atp9*), Apocytochrom-b (*cob*), drei Cytochrom-c-Oxidase-Untereinheiten (*cox1*, *cox2*, *cox3*) und sieben NADH-Dehydrogenase-Untereinheiten (*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*). Alles typische mitochondriale Gene, die man auch in anderen Pilzen findet (Aguileta et al., 2014). Zusätzlich fanden sich in den mtDNAs der betrachteten Zygomyceten Gene für die kleine und die große rRNA-Untereinheit (*rns*, *rnl*), sowie für eine Ribonuklease-P-RNA (*rnpB*).

Backusella circina und *Lichtheimia corymbifera* bilden die Ausnahme. In *Backusella circina* fehlen das *nad6*-Gen und einige tRNAs. Da die mtDNA von *Backusella circina* in

der JGI MycoCosm-Datenbank nicht annotiert ist, ist es möglich, dass die mitochondriale Sequenz dort unvollständig ist. Ein Teil des *nad6*-Gens wurde auf einem kurzen Scaffold (scf884) zusammen mit einer vollständigen Kopie des *nad1*-Gens gefunden. Die Sequenz der mtDNA lässt sich mit diesem Scaffold verlängern. Ein unvollständig annotierter Scaffold wurde als mtDNA von *Lichtheimia corymbifera* (NCBI: CBTN010000097) identifiziert. Der Datenbankeintrag bezieht sich auf einen linearen Contig (contig SCAF00098), der einzelne mitochondriale Gene enthält. Dieser Contig hat eine Länge von 83 533 bp. Eine Duplikation von mtDNA in Zygomyceten ist nicht bekannt. Die lineare Sequenz lässt sich mit gerade einmal vier Deletionen innerhalb von Bereichen mit einfachen Wiederholungen (CCCCCC, CCCCCC, AAAA, ATATAT) in ein ringförmiges Molekül umwandeln. Die kürzere geschlossene mtDNA-Sequenz mit einer Länge von 43 132 bp wurde für alle weiteren Analysen benutzt.

Die Interpretationen der in den Datenbanken nicht annotierten mtDNAs von *Absidia repens*, *Backusella circina*, *Lichtheimia corymbifera*, *Parasitella parasitica* NRRL 2501, *Rhizopus microsporus* und *Umbelopsis ramanniana*, sowie der beiden sequenzierten Zygomyceten *Absidia glauca* und *Parasitella parasitica* CBS 412.66 befinden sich als GenBank-Dateien auf dem zu dieser Arbeit gehörenden Datenträger.

Der Vergleich von Chondriomen verschiedener Zygomyceten zeigt, dass es in *Parasitella parasitica* eine große Anzahl an Endonukleasen gibt. Im Allgemeinen lassen sich die mtDNAs von Pilzen mit einem konservierten Genbestand und einer stark variierenden Genfolge charakterisieren (Aguileta et al., 2014). In den meisten Ascomyceten werden Gene nur auf einem Strang der mtDNA codiert. In Basidiomyceten befinden sich Gene auf beiden Strängen der mtDNA. Die mtDNAs von Pilzen beinhalten große Bereiche in und zwischen den Genen, die Introns, meist aus Gruppe I, und repetitive Elemente beinhalten. Die Anzahl der Introns variiert stark. Es gibt Pilze ohne Introns und andere, die mehr als zehn Introns in ihrer mtDNA haben.

		Absgl	Bacci	Glosp	Licco	Licra	Morve	Mucci	Parpa	Phybl	Rhimi	Rhior	Umbra	Zancu
Länge (bp)		63 080	>33 873	86 170	43 132	31 830	58 745	31 831	83 361	62 082	>34 812	54 178	>30 438	58 654
GC	Min	13 %	23 %	28 %	16 %	26 %	20 %	23 %	20 %	24 %	15 %	12 %	25 %	10 %
	Ø	28 %	34 %	37 %	26 %	34 %	28 %	34 %	31 %	38 %	25 %	26 %	33 %	19 %
	Max	46 %	45 %	47 %	44 %	45 %	37 %	45 %	50 %	57 %	44 %	46 %	47 %	31 %
Gene und Introns														
<i>atp6</i>		+	+	+	+	+	+	+	+	+	+	+	+	+
<i>atp8</i>		+	+	+	+	+	+	+	+	+	+	+	+	+
<i>atp9</i>		+ (1)	+	+	+	+	+	+	+ (3)	+	+ (1)	+ (1)	+	+
<i>cob</i>		+ (1)	+ (1)	+ (4)	+ (1)	+	+	+	+ (7)	+ (3)	+	+ (2)	+ (1)	+ (3)
<i>cox1</i>		+ (5)	+	+ (11)	+	+ (1)	+ (3)	+ (1)	+ (8)	+ (8)	+ (3)	+ (3)	+	+ (9)
<i>cox2</i>		+ (1)	+	+	+	+ (1)	+	+	+ (1)	+	+ (1)	+ (1)	+	+ (1)
<i>cox3</i>		+ (3)	+	+ (3)	+ (1)	+	+ (1)	+	+ (3)	+	+ (1)	+ (1)	+	+
<i>dpo</i>		+	-	+	-	-	-	-	+	-	-	-	-	+
<i>nad1</i>		+ (3)	+	+	+	+	+	+	+ (1)	+ (3)	+	+	+	+
<i>nad2</i>		+	+	+	+	+	+	+	+	+	+	+	+	+
<i>nad3</i>		+ (2)	+	+	+	+	+	+	+	+	+	+ (1)	+	+
<i>nad4</i>		+ (1)	+	+	+	+	+	+	+	+	+	+	+	+
<i>nad4L</i>		+	+	+	+	+	+	+	+	+	+	+	+	+
<i>nad5</i>		+ (1)	+	+ (2)	+	+	+	+	+ (1)	+ (1)	+	+	+	+
<i>nad6</i>		+ (1)	(-)	+	+	+	+	+	+	+	+	+	+	+
<i>rns</i>		+	+	+	+	+	+	+	+	+	+	+	+	+
<i>rnl</i>		+	+	+ (5)	+ (2)	+	+	+	+ (3)	+ (2)	+	+	+	+ (1)
<i>rnpB</i>		+	+	-	+	+	+	+	+	+	+	+	+	+
<i>rps3</i>		-*	-*	-	-*	-	+	-*	-*	-*	-*	-*	-*	+
Endonukleasen (innerhalb von Introns)														
GIY-YIG		8 (5)	1 (1)	1	-	-	5 (2)	-	7 (5)	3 (3)	1 (1)	2 (1)	2 (1)	2 (2)
LAGLIDADG		4 (4)	-	11 (8)	1 (1)	-	2 (1)	-	19 (18)	9 (9)	2 (2)	4 (4)	-	11 (11)
Hypothetische Proteine (innerhalb von Introns)														
ORF		1	1	5 (1)	3	1	2	1	-	5	-	3	1	2
Zweite Kopie		<i>nad2</i>	-	-	-	-	<i>cox2</i>	-	-	-	-	<i>atp9</i>	-	-
tRNAs														
Anzahl		24	>20	25	23	25	26	24	26	24	25	24	>16	27

Tabelle 3.1: Übersicht der Gene in mtDNAs von Zygomyceten und *Glomus sp.*

GC: Zahlen zeigen den GC-Gehalt der mtDNAs, berechnet mit GC-Draw (siehe Abbildung 3.6) und einer Fenstergröße von 1000 nt.

+: Das Gen existiert in diesem Pilz.

-: Das Gen existiert nicht in diesem Pilz.

(-): Das Gen existiert nicht in der mtDNA von *Backusella circina*, wurde aber als Fragment auf einem anderen Scaffold gefunden, mit dem sich die mtDNA verlängern lässt.

*: Das Gen befindet sich im Kerngenom.

>: Die Länge der mtDNA und die Anzahl der tRNAs wurde unterschätzt, da die mtDNA unvollständig ist.

Die Zahlen in Klammern stehen für die Anzahl an Introns oder sich in Introns befindenden Elementen.

(Tabelle nach Ellenberger, Burmester und Wöstemeyer (2016b))

Die Länge der Introns schwankt stark, zwischen einzelnen bp und bis zu 5 kb. Ein anderes unterscheidbares Merkmal der mtDNAs von Pilzen ist die Verteilung von tRNAs. Sie sind entweder über das gesamte Chondriom verteilt oder finden sich zu einigen wenigen Gruppen zusammen.

Tabelle 3.1 zeigt den Gengehalt der untersuchten mtDNAs von Zygomyceten. Abbildung 3.2 zeigt die lineare Genomrepräsentation der mtDNAs von Zygomyceten zum Vergleich der mitochondrialen Synthese. Die Größe der Chondriome schwankt zwischen rund 32 kb in *Mucor circinelloides* und 62 kb in *Phycomyces blakesleeanus*. *Absidia glauca* und *Parasitella parasitica* CBS 412.66 gehören zu den Pilzen mit größeren Chondriomen. Ihre mtDNAs haben eine Größe von rund 63 kb und 83 kb. In diesem Vergleich war nur die mtDNA von *Glomus sp.* mit über 86 kb noch größer. Das Auftreten mehrerer DNA-Polymerasen (*dpo*) erzeugt diese ungewöhnliche Sequenzlänge.

Abbildung 3.3 zeigt die lineare Genomrepräsentation der unvollständig assemblierten mtDNAs einer zweiten *Parasitella parasitica*, *Rhizopus microsporus* und *Umbelopsis ramaniana*. In den mtDNAs von Zygomyceten werden Gene typischerweise auf beiden DNA-Strängen codiert. In *Glomus sp.* hingegen befinden sich nur die DNA-Polymerase-Gene und einige offene Leserahmen auf dem gegenüberliegenden Strang. Obwohl der Gengehalt größtenteils konserviert ist (Tabelle 3.1), ist die Anordnung der Gene sehr unterschiedlich.

In vielen der betrachteten Pilze finden sich kleine Gruppierungen aus zwei dicht aufeinander folgenden Genen (Tabelle 3.2). Es kommt sogar vor, dass sich die codierenden Bereiche zweier Gene (*nad2-nad3* und *nad4L-nad5*) an einem Nukleotid überschneiden. Das bedeutet, dass das letzte Nukleotid des Stopcodons UAA des einen Gens gleichzeitig das erste Nukleotid des Startcodons AUG des zweiten Gens ist. Diese Überschneidungen findet man zum Beispiel in *Phycomyces blakesleeanus* und *Rhizopus oryzae*.

Zusätzlich zu den typischen Genen beinhalten die mtDNAs aller betrachteten Zygomyceten das *rnpB*-Gen. Dieses Gen codiert die RNA-Untereinheit der mitochondrialen RNase-P. Diese bildet den enzymatisch aktiven Teil einer Endonuklease (Ribonukleoprotein), die für die Reifung der tRNA verantwortlich ist. Die Größe und Sequenz der RNA-Untereinheit variiert erheblich, was ihre Identifikation deutlich erschwert. Das Gen ist weder in den mtDNAs von Basidiomyceten, noch in denen von Chytridiomyceten zu finden (Seif et al., 2003). In Ascomyceten tritt es nur ungleichmäßig verteilt auf. In Zygomyceten wurde das *rnpB*-Gen mit einer Mustersuche nach den konservierten Regionen CRI (GGAAAGTC) und CRV (ACANAANNNGCTTAT) (Chen und Pace, 1997; Seif et al., 2003) gefunden. Die Lage dieses Gens ist in den meisten Zygomyceten konserviert. In der mtDNA aller betrachteten Zygomyceten, mit Ausnahme von *Umbelopsis ramaniana* und *Zancudomyces culisetae*, wird *rnpB* von zwei tRNAs flankiert.

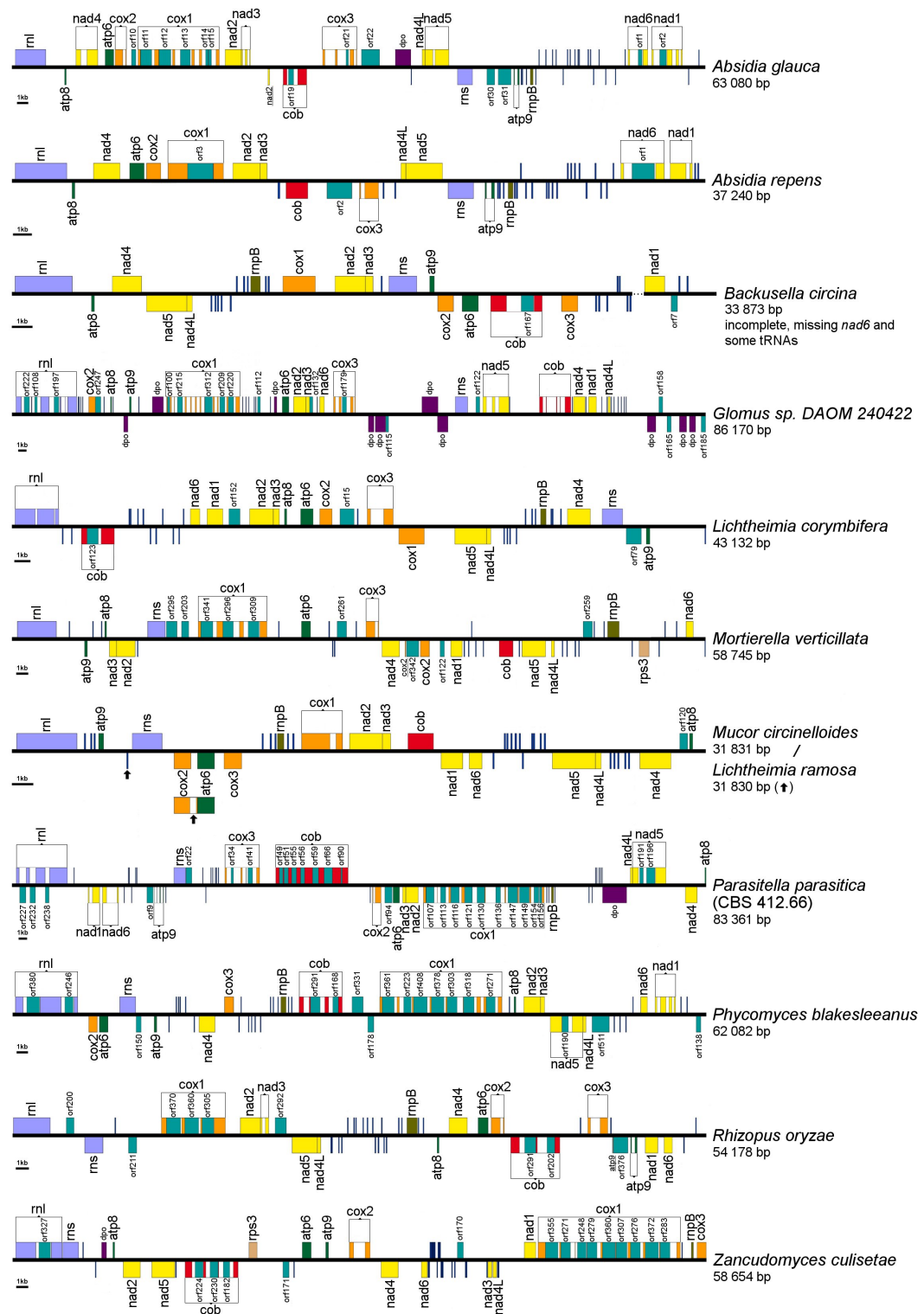


Abbildung 3.2: Lineare Genomrepräsentation zum Vergleich der Synthenie in mtDNAs verschiedener Zygomyceten und dem arbuskulären Mykorrhizapilz *Glomus sp.* Die linearen mtDNAs wurden oberhalb des *rnl*-Gens geöffnet, um einen Vergleich zu erleichtern. Fragmente einer zweiten Kopie eines mitochondrialen Gens sind unterstrichen. Die Pfeile deuten auf Unterschiede zwischen den nahezu identischen mtDNAs von *Mucor circinelloides* und *Lichtheimia ramosa* hin. Weiße Bereiche innerhalb eines Gens sind Introns. Die dünnen blauen Balken sind tRNAs. Die mtDNA von *Backusella circina* ist unvollständig. Das Gen für *nad6* fehlt. Es wurde auf einem anderen Scaffold, der nicht als Teil der mtDNA gekennzeichnet ist, gefunden. „...“ bezeichnet die Stelle, an der sich die mtDNA von *Backusella circina* mit einem zusätzlichen Scaffold verlängern lässt. (Abbildung nach Ellenberger, Burmester und Wöstemeyer (2016b))

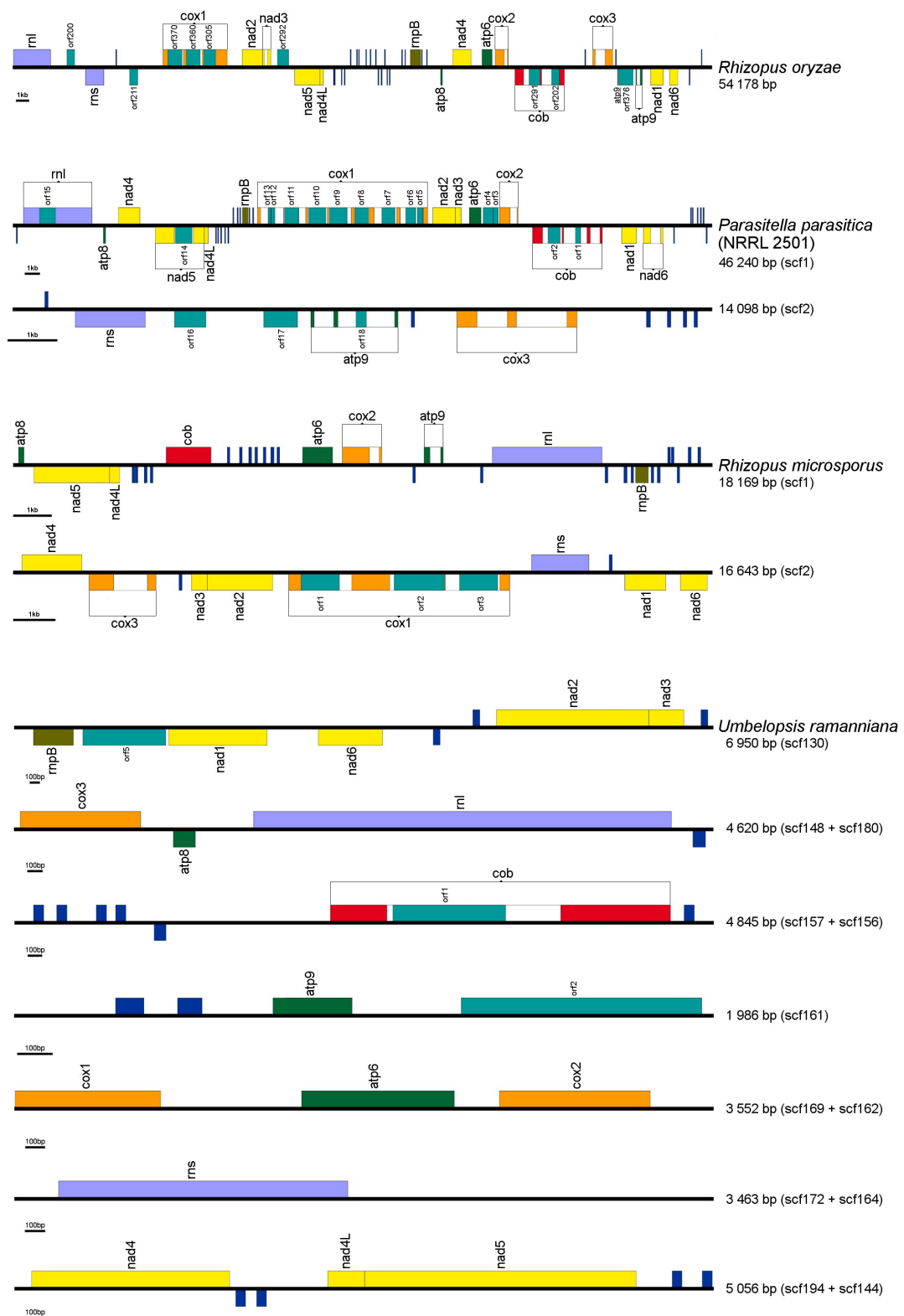


Abbildung 3.3: Lineare Genomrepräsentation zum Vergleich der Synthese in den unvollständigen mtDNAs einer zweiten *Parasitella parasitica*, *Rhizopus microsporus* und *Umbelopsis ramanniana*. Die mtDNA von *Rhizopus oryzae* dient dem Vergleich zu einer vollständig assemblierten mtDNA. Weiße Bereiche innerhalb eines Gens sind Introns. Die dünnen blauen Balken sind tRNAs. (Abbildung nach Ellenberger, Burmester und Wöstemeyer (2016b))

	Absgl	Bacci	Glosp	Licco	Licra	Morve	Mucci	Parpa	Phybl	Rhimi	Rhior	Umbra	Zancu
<i>atp6-cox2</i>	■	■		■	■		■	■	■	■	■	■	
<i>atp8-nad4</i>	■	■			■		■	■		■	?	?	
<i>cob-cox3</i>	■	■						■	■		■	?	
<i>nad1-nad6</i>	■	?		■	■		■	■	■	■	■	■	
<i>nad2-nad3</i>	■	■	■	■	■	■	■	■	■	■	■	■	
<i>nad4L-nad5</i>	■	■		■	■		■	■	■	■	■	■	

Tabelle 3.2: Übersicht der Gene, die in mtDNAs von Zygomyceten oft nebeneinander liegen.

■: Gene, die paarweise auftreten

? : hypothetisches Genpaar

Da die mtDNAs von *Rhizopus microsporus* und *Umbelopsis ramanniana* nicht vollständig assembliert sind, fehlen verbindende Teilsequenzen zwischen den beiden Genen der Genpaare *atp8-nad4* und *cob-cox3*. Die mtDNA von *Backusella circina* lässt sich mit einem Scaffold erweitern, auf dem sich ein *nad6*-Fragment und ein vollständiges *nad1*-Gen befinden. Daher ist auch das Genpaar *nad1-nad6* in *Backusella circina* hypothetisch. (Tabelle entnommen aus Ellenberger, Burmester und Wöstemeyer (2016b))

Meist sind das zwei Kopien der Methionin-tRNA. In *Glomus intraradices* fehlt ein vollständiges *rnpB*-Gen (Lee und Young, 2009). In *Glomus sp.* wurde ebenfalls kein zusammenhängendes *rnpB*-Gen gefunden. Allerdings sind die beiden konservierten Regionen an unabhängigen Positionen der mtDNA vorhanden. CRI startet bei Position 72 796 und CRV bei Position 39 603 der mtDNA. Abbildung 3.4 zeigt ein lokales Alignment der konservierten Regionen CRI und CRV der *rnpB*-Sequenzen aus Zygomyceten, *Glomus sp.* und anderen Pilzen. Eine neue konservierte Region (CR zygo) wurde direkt nach CRV lokalisiert. Das Muster TCTAAGA scheint für Zygomyceten charakteristisch zu sein. Zusätzlich betrachtete *rnpB*-Gene aus anderen Pilzen sind kürzer oder beinhalten nicht dieses spezifische Sequenzmotiv.

Absidia glauca, *Parasitella parasitica* und *Glomus sp.* besitzen zusätzlich zu den bereits besprochenen Genen DNA-Polymerase-Gene (*dpo*) in ihren Chondriomen. In den mtDNAs von *Absidia glauca* und *Parasitella parasitica* CBS 412.66 gibt es jeweils nur ein *dpo*-Gen. *Glomus sp.* besitzt zehn *dpo*-Gene. Sie sind der Grund für die besondere Größe dieser mtDNA.

Ein weiteres Unterscheidungsmerkmal zwischen den mtDNAs von Zygomyceten ist das Gen für das Protein der kleinen ribosomalen Untereinheit 3 (*rps3*). Es findet sich nur in den mtDNAs von *Mortierella verticillata* und *Zancudomyces culisetiae*. Seif et al. (2005) gehen davon aus, dass die *rps3*-Gene innerhalb der Opisthokonta drei mal unabhängig voneinander verlorengegangen sind. Sie nehmen an, dass in allen Fällen diese Gene in das Kerngenom transferiert wurden, genau wie die Gene anderer ribosomaler Proteine, die in den mitochondrialen Genomen von Pilzen fehlen (Lang et al., 1999).

<i>A. glauca</i>	42	GAGGAAAGUC	CGGGGCU	58 ...	225	AUUACAAA	AUCCGGCUUA	AUAUAU	UCUAAGA	255
<i>B. circina</i>	41	GAGGAAAGUC	CGAGGCU	57 ...	440	AUAACAGAA	AUCCGGCUUA	UUUAGU	UCUAAGA	470
<i>L. corymbifera</i>	48	UGGAAAGUC	CGAGAGC	64 ...	326	AUAACUAAA	AUCCGGCUUA	UCAAGU	UCUAAGA	356
<i>L. ramosa</i>	44	GAGGAAAGUC	CGGGGCU	60 ...	261	AUAACAAA	AUCCGGCUUA	UUUAUC	UCUAAGA	291
<i>M. verticillata</i>	17	AAGGAAAGUC	CAGAAUA	33 ...	942	UGUACUGAA	UCUGGCUUA	AUAUAA	UCUAAGA	972
<i>M. circinelloides</i>	44	GAGGAAAGUC	CGGGGCU	60 ...	261	AUAACAAA	AUCCGGCUUA	UUUAUC	UCUAAGA	291
<i>P. parasitica</i>	18	GAGGAAAGUC	CGGGGCU	34 ...	235	AUAACAAA	AUCCGGCUUA	UUUAUC	UCUAAGA	265
<i>P. blakesleeanus</i>	44	GAGGAAAGUC	CGGGGCU	60 ...	440	AUUACAGAA	AUCCGGCUUA	UAUAGU	UCUAAGA	470
<i>R. microsporus</i>	41	AAGGAAAGUC	CGGAGCU	57 ...	300	AUAACAUAA	AUCCGGCUUA	UUUAAU	UCUAAGA	330
<i>R. oryzae</i>	42	GAGGAAAGUC	CGGGGCU	58 ...	789	AUAACAGAA	AUCCGGCUUA	UUUAAC	UCUAAGA	819
<i>U. ramanniana</i>	13	AAGGAAAGUC	CGAGGUU	29 ...	286	AUUACAAA	AUCCGGCUUA	UCUAAU	UCUAAGU	316
<i>Z. culisetae</i>	14	UAGGAAAGUC	CGAAUUA	30 ...	145	UUUACAGAA	AUCCGGCUUA	UAUUUA	AAAUUCU	175
<i>Glomus sp.</i>	X	UAGGAAAGUC	CGGGGCU	X ...	X	CUUACAUAA	CCCCGGCUUA	UAUGGAA	CUAGAG	X
<i>S. cerevisiae</i>	19	UAGGAAAGUC	AUAAUUA	35 ...	448	UAUACAGAA	AUAUUGCUUA	AUUUAU	AAUAUAAU	478
<i>E. coli</i>	61	GAGGAAAGUC	CGGGGCU	77 ...	344	ACGACAGAA	CCCCGGCUUA	UCGGUCAG	UUUCA	374
		CR I				CR V			CR zygo	

Abbildung 3.4: Lokales Alignment konservierter Regionen des *mpB*-Gens in Zygomyceten.

In *Glomus sp.* wurden die konservierten Regionen CRI und CRV gefunden. Sie ergeben jedoch kein zusammenhängendes Gen, sondern befinden sich in der mtDNA an unterschiedlichen Positionen (X). Mit CR zygo (rot markiert) wurde eine für Zygomyceten einzigartige konservierte Region definiert. Universell konservierte Nukleotide sind fett gedruckt (Chen und Pace, 1997). Grau hinterlegte Nukleotide sind zu über 75 % konserviert (Seif et al., 2003). Die Zahlen bezeichnen die Sequenzposition im lokalen Alignment. (Abbildung entnommen aus Ellenberger, Burmester und Wöstemeyer (2016b))

Tatsächlich wurden die fehlenden *rps3*-Gene in den Kerngenomen aller in dieser Arbeit betrachteten Pilze identifiziert, mit Ausnahme von *Glomus sp.* und *Lichtheimia ramosa*, bei denen nur die mitochondrialen Genome verfügbar sind. Abbildung 3.5 zeigt ein Alignment der *rps3*-Sequenzen.

Die Anzahl an Introns in mtDNAs von Pilzen ist sehr unterschiedlich. Die höchste Anzahl an Introns gibt es in *Parasitella parasitica* CBS 412.66 mit 27 Introns. Die meisten davon gehören zur Gruppe I und beinhalten Endonukleasen. Andere Zygomyceten, wie *Backusella circina*, *Mucor circinelloides* und *Lichtheimia ramosa*, haben nur ein oder zwei Introns in ihren mtDNAs und besitzen keine Endonukleasen. Endonukleasen in Zygomyceten sind vom GIY-YIG- und LAGLIDADG-Typ und sind sehr unterschiedlich verteilt. Generell kommen Endonukleasen vom LAGLIDADG-Typ etwa zweimal so häufig in Zygomyceten vor, wie solche vom GIY-YIG-Typ. Die Sequenzidentität der verschiedenen Endonukleasen ist vergleichbar, wenn man die LAGLIDADG-Endonukleasen nach ihrer Position gruppiert und nur innerhalb dieser Gruppen paarweise vergleicht. Im Durchschnitt ergibt sich eine Sequenzidentität von rund 25 %. Tabelle 3.3 zeigt die Verteilung von Endonukleasen in mitochondrialen Genen von Zygomyceten. Anzahl und Typ der Endonukleasen variieren stark. Besonders *cob* und *cox1* sind Reservoirs für Endonukleasen. Bis zu 50 % der Endonukleasen zeigen Ähnlichkeiten zu Endonukleasen aus *Parasitella parasitica* CBS 412.66. Endonukleasen mit höherer Sequenzähnlichkeit findet man sehr oft innerhalb des gleichen Gens. Besonders hoch sind die Sequenzähnlichkeiten bei Endonukleasen in *cox1*.

```

Smicu_rps3      MGKRHSTSLKLRHIGNWNNNISYNNNN-----INKEKIKFILNSYFIEKAIISEPIFI
Morve_rps3      MGSKHPLSLRLKTHNKWGETSYFNSKSELHNYWGVDKIALNVIKH-YFQFALISKPKFN
Umbra_rps3      -MSKSVKALGLKQKQDWETSVCFSRNNRNNLTWEKDQAARTTLKH-YFQKHLTSTPHFK
Licco_rps3      -----MWEKDGAAKEALKH-FFQNQLTSTPQFR
Phybl_rps3      -MSKSIKALS LK GKQDWNTSIYSFTRNVRNNLLWEKDGA AKQSLKV-YFQQYLTSTPAFR
Absgl_rps3      -MSKSIGALGLK GKQDWKTSIYSFTRNTRNNILWEKDGS AKTTLKH-FFQSHLTSTPTFH
Rhimi_rps3      -MSKSIKALS LK GKQDWDTSVFCFSRNVNRMMLWEKDAAAKATLKH-FFQKHLTSTPAFK
Rhior_rps3      -----MLWEKDAAAKATLKH-FFQKHLTSTPAFK
Bacci_rps3      -MSKSIKALS LK GKQEWNTSVYCF SRNARNMMWEKD GASKAALKH-FFQNHLTSTPAFK
Mucci_rps3      -MSKSIKALS LK GKQDWDTSVYCF SRNARNMMMLWEKD GATKATLKH-FFQKHLTSTPAFR
Parpa_rps3      -MSKSIKALS LK GKQDWDTSVYCF SRNTRNNILWEKD GATKTTLKH-FFQKHLTSTPAIR
                :          ::   :::   : * * :

Smicu_rps3      YKLNEIKIIIIYWSSFNLLLD-----KDINKILLY
Morve_rps3      KDSNKIVISFYFYLNL PVS KRFPY GQSKRIQSS LQGDVTANNK NLLPLEAKGYSKLILK
Umbra_rps3      HSTAKVTVQMYYYADPNLVTL-----SEFEIKPLE-----QM--L
Licco_rps3      HSPSRVTIQMFYYADPNFGDL-----SQFKLQPLE-----QL--V
Phybl_rps3      HAPSRVTIQMFYYADPNFSNL-----QNFKLEPLE-----KL--L
Absgl_rps3      HSPLKVS IQLFYYTDP SVGRA-----ENFKLEPLE-----KL--L
Rhimi_rps3      HSPSRVTIQLFYYVDP AFGNA-----SSFKLDAL E-----KL--I
Rhior_rps3      HSPSRVTIQLFYYADPEFGNA-----SSFKLEPLE-----KL--I
Bacci_rps3      HSPSKVTIQLFYYTDSL V-NA-----AEFKLEPLE-----KL--I
Mucci_rps3      HSPSKVTIQLYYYADPAFGHV-----ADFKLEPLE-----KL--I
Parpa_rps3      HSPSKVTIQLYYYADPILGRA-----ADFKLEPLE-----KL--I
                .: : ::*: . . .                               ::

Smicu_rps3      INNILNIKVNINIKLKKPYLSSKILSEYISINIKKYNIKKIAKKIIKQIKIEKKLLNPL
Morve_rps3      LSRFYGKPVLELRPVRIHYPYLSYILAQYIAINIRLGNFNKLMRTLFKKAKLVRNNNINT
Umbra_rps3      QQIYPQKYVDVRVRLHYPYLN AEILAKYLTINAQRSSWSVLTRKFMRGVP IVKPPLPHH
Licco_rps3      SNLYPNKYVDMRVIRLHYPYMN AEILAKYLTVNSQKKSWSALTRPFMKNM PIVKAPLAHN
Phybl_rps3      STMPYNKFVDLRFVRVHYPYMN AEILAQYLTVNAERSGWSMLTRKFMKGVP LV-----
Absgl_rps3      RSMYPNKYVDLRLIRVHYPYMN AEILAQYLTDNAERSGWS S LTRKFMKGVP MVK PPLPHN
Rhimi_rps3      GSFYPNKFVDLRLTQLHYPYMN AEILAQFLTINAERNGWSMVNRRFFKAAP IVKAPLPHN
Rhior_rps3      GSFYPNKFIDLRLTQVHYPYMN AEILAQYLTINAERNGWSAVNRRFLKAAP IVKAPLPHN
Bacci_rps3      SSLYPNKYVDLRFVKVHYPYMN AEILAQYLTINSQRVGWSQLTRKFMRGVP MVK PPLPHN
Mucci_rps3      TSLYPGKFVDLRFVQVHYPYMN AEILAQYLTINSERIGWSQLTRRFMKGVP LVK PPLPHN
Parpa_rps3      TSLYPGKFVDLRFVQVHYPYMN AEILAQYLTINAERMGWSQLTRRFMKGVP LVK PPLPHN
                : : .   : : * : : : * : : : : * . . : : : : :

Smicu_rps3      I--NFNYYNIIKYYYINKKLYSSISGIKLVFKGRLSKRRTAERKKILTFNKGFLNKNLSLFT
Morve_rps3      T----NFRDLIKYSSLINRPQFLSGLKIQISGRLSQRKAASRTIRVRSIGTLRLSSSTS
Umbra_rps3      MPPAVDWN NLI PHASNGQLTSAIQGVKYQVSGRLGRRRGAGRSSILR KSLGTFKFQTTAKS
Licco_rps3      MPFAVQWQNLLPQVANDKLTTAIQGVKYQVSGRLGRRSGSGRSQILRKSIGTFQFTSHRS
Phybl_rps3      -----ANGQLTSAIQGIKYQVSGRLGRRKGASRTQVMR KSLGTFQFTSHKS
Absgl_rps3      MPFGVQWNNLLPQVGNAQLTSAIQGVKYEVSGRLGRRKGASRTQVLR KSLGTFQFTSHKS
Rhimi_rps3      MPFATQWTSLLPQLSNGQLTSAIQGIKYQVSGRLGRRKGASRTQVMR KSLGTFQFTSHKS
Rhior_rps3      MPFAVQWNSLLPQVSNQQLTSAIQGIKYQVSGRLGRRKGASRTQVMR KSVGTFQFTSHKS
Bacci_rps3      MPFGVQWNSLLPQVSNQQLTSAIQGIKYQVSGRLGRRKGASRTQVMR KSLGTFQFTSHKS
Mucci_rps3      MPFAVQWNSLLPQVSNQQLTSAIQGIKYEVSGRLGRRKGASRTQVMR KSLGTFQFTSHKS
Parpa_rps3      MPFAVQWHSLLPQVSNQQLTSAIQGIKYEVSGRLGRRKGASRTQVLR KSLGTFQFTSHKS
                :          : : * : * . * * : * : * . : : : :

Smicu_rps3      LIDYKSPYNFKFINGSIGIKTILNNTIHF-----
Morve_rps3      LIDA-SKFSFKGNGAATVKVWLSSCTVNP SVNKNTLINTP-----P-LHSTLII
Umbra_rps3      LIDV-GRHTFANRNGSITVKVWISSALFGATALNKASQKAATIKKP-----
Licco_rps3      LVDV-GRHTFSNKNGSITVKVWIASALFGANALSRHV NKKAVESAAAAAPT PKL-
Phybl_rps3      LVDV-GRHTFANKNGSITVKVWIASALFGVNALAKKITQKAAESAKL-----
Absgl_rps3      LVDV-GRHTFANKNGSITVKVWIASALFGVNALAKKMASSSNVGAVA AKK-----
Rhimi_rps3      LVDV-GRHTFANKNGSITVKVWIATALFGVNTAVKNISKM-----
Rhior_rps3      LVDV-GRHTFANKNGSITVKVWIASALFGVNALAKSAGKKVIESAKL-----
Bacci_rps3      LVDV-GRHTFANKNGSITVKVWISSALFGVGALANKMGRKAVESAKL-----
Mucci_rps3      LVDV-GRHTFANKNGSITVKVWIASALFGVNAMTK-LGKKALEKSAAAAAKL----
Parpa_rps3      LVDV-GRHSFANKNGSITVKVWIASALFGVNAMAR-LGKKALEKGAPAAASSAKL
                * : * . . *   ** : : * : .

```

Abbildung 3.5: Lokales Alignment einiger *rps3*-Gene aus Zygomyceten.
(Abbildung entnommen aus Ellenberger, Burmester und Wöstemeyer (2016b))

	Absgl	Bacci	Glosp	Licco	Licra	Morve	Mucci	Parpa	Phybl	Rhimi	Rhior	Umbra	Zancu
<i>cob</i>	1xL	1xG		1xL				1xG 6xL	2xL		1xG 1xL	1xG	1xG 2xL
<i>cox1</i>	3xG 2xL		4xL			2xG 1xL		4xG 6xL	3xG 4xL	1xG, 2xL	3xL		1xG 8xL
<i>cox3</i>	1xL		1xL					1xL					
<i>nad1</i>	1xG												
<i>nad5</i>								2xL	1xL				
<i>nad6</i>	1xG												
<i>rnl</i>			3xL					3xL	2xL				1xL
außerhalb von Genen	3xG		1xG 3xL			3xG 1xL		2xG 1xL			1xG	1xG	

Tabelle 3.3: Übersicht der Art und Lage von Endonukleasen in mtDNAs von Zygomyceten.

Parasitella parasitica CBS 412.66 besitzt die meisten Endonukleasen. Die Gene für *cob* und *cox1* enthalten besonders oft Introns, die für Endonukleasen codieren.

G: GIY-YIG-Endonuklease

L: LAGLIDADG-Endonuklease

(Tabelle nach Ellenberger, Burmester und Wöstemeyer (2016b))

Tabelle 3.4 zeigt, wie unterschiedlich mtDNAs selbst bei nah verwandten Pilzen sind. Bei *Absidia repens* finden sich nur drei Endonukleasen innerhalb der mtDNA, im Gegensatz zu 12 Endonukleasen bei *Absidia glauca*. Die mtDNA von *Absidia repens* beinhaltet nur fünf Introns, deutlich weniger als bei anderen Zygomyceten. Die Anordnung der mitochondrialen Gene ist bei den beiden *Absidia*-Stämmen nahezu gleich (Abbildung 3.1, Abweichung bei der Orientierung von *cox3*). Der GC-Gehalt der mtDNA ist bei *Absidia repens* geringer als bei *Absidia glauca*. Abbildung 3.6 zeigt den GC-Gehalt der beiden *Absidia*-Stämme im Vergleich als ringförmige Darstellung, die mit GC-Draw (Abschnitt 6.1) erstellt wurde.

Die Anzahl und Verteilung von Endonukleasen in zwei Stämmen von *Parasitella parasitica* ist ebenfalls unterschiedlich. *Parasitella parasitica* CBS 412.66 hat 26 Endonukleasen innerhalb der mtDNA. Davon sind 19 Endonukleasen vom LAGLIDADG-Typ. In der mtDNA von *Parasitella parasitica* NRRL 2501 finden sich 18 Endonukleasen. Hier sind die meisten allerdings vom GIY-YIG-Typ (12 Endonukleasen). *Parasitella parasitica* CBS 412.66 hat 27 Introns in der mtDNA, *Parasitella parasitica* NRRL 2501 besitzt 18 Introns. Die Anordnung der mitochondrialen Gene ist bei den beiden *Parasitella*-Stämmen im Bereich von *rnl* bis *cob* gleich (Abbildungen 3.1, 3.4). Die Anordnung der Gene auf dem zweiten mitochondrialen Scaffold von *Parasitella parasitica* NRRL 2501 unterscheidet sich von der in *Parasitella parasitica* CBS 412.66 (*rns-atp9-cox3* vs. *atp9-rns-cox3*). Der GC-Gehalt der mtDNA ist bei *Parasitella parasitica* NRRL 2501 höher als bei *Parasitella parasitica* CBS 412.66. Abbildung 3.6 zeigt den GC-Gehalt der beiden *Parasitella*-Stämme im Vergleich als ringförmige Darstellung, die mit GC-Draw (Abschnitt 6.1) erstellt wurde.

		Absgl (CBS 101.48) (-)-Kreuzungstyp	Absre (NRRL 1336) (+)-Kreuzungstyp	Parpa (CBS 412.66) (+)-Kreuzungstyp	Parpa (NRRL 2501) (-)-Kreuzungstyp
Länge (bp)		63 080	37 240	83 361	>60 338
GC	Min	13 %	14 %	20 %	46 %
	Ø	28 %	26 %	31 %	35 %
	Max	46 %	35 %	50 %	23 %
Gene und Introns					
<i>atp6</i>		+	+	+	+
<i>atp8</i>		+	+	+	+
<i>atp9</i>		+ (1)	+ (1)	+ (3)	+ (2)
<i>cob</i>		+ (1)	+	+ (7)	+ (3)
<i>cox1</i>		+ (5)	+ (1)	+ (8)	+ (7)
<i>cox2</i>		+ (1)	+	+ (1)	+ (1)
<i>cox3</i>		+ (3)	+ (1)	+ (3)	+ (2)
<i>dpo</i>		+	-	+	-
<i>nad1</i>		+ (3)	+ (1)	+ (1)	+
<i>nad2</i>		+	+	+	+
<i>nad3</i>		+ (2)	+	+	+
<i>nad4</i>		+ (1)	+	+	+
<i>nad4L</i>		+	+	+	+
<i>nad5</i>		+ (1)	+	+ (1)	+ (1)
<i>nad6</i>		+ (1)	+ (1)	+	+ (1)
<i>rns</i>		+	+	+	-
<i>rnl</i>		+	+	+ (3)	+ (1)
<i>rnpB</i>		+	+	+	+
<i>rps3</i>		_*	_*	_*	_*
Endonukleasen (innerhalb von Introns)					
GIY-YIG		8 (5)	3 (2)	7 (5)	12 (8)
LAGLIDADG		4 (4)	-	19 (18)	6 (6)
Hypothetische Proteine (innerhalb von Introns)					
ORF		1	-	-	-
Zweite Kopie		<i>nad2</i>	-	-	-
tRNAs					
Anzahl		24	26	26	24

Tabelle 3.4: Vergleich der Gene in mtDNAs von *Absidia glauca* und *Absidia repens*, sowie zwei Stämmen von *Parasitella parasitica* mit unterschiedlichem Kreuzungstyp.

GC: Zahlen zeigen den GC-Gehalt der mtDNAs, berechnet mit GC-Draw (siehe Abbildung 3.6) und einer Fenstergröße von 1000 nt.

+: Das Gen existiert in diesem Pilz.

-: Das Gen existiert nicht in diesem Pilz.

_: Das Gen befindet sich im Kerngenom.

>: Die Länge der mtDNA und die Anzahl der tRNAs wurde unterschätzt, da die mtDNA unvollständig ist.

Die Zahlen in Klammern stehen für die Anzahl an Introns oder sich in Introns befindenden Elementen.

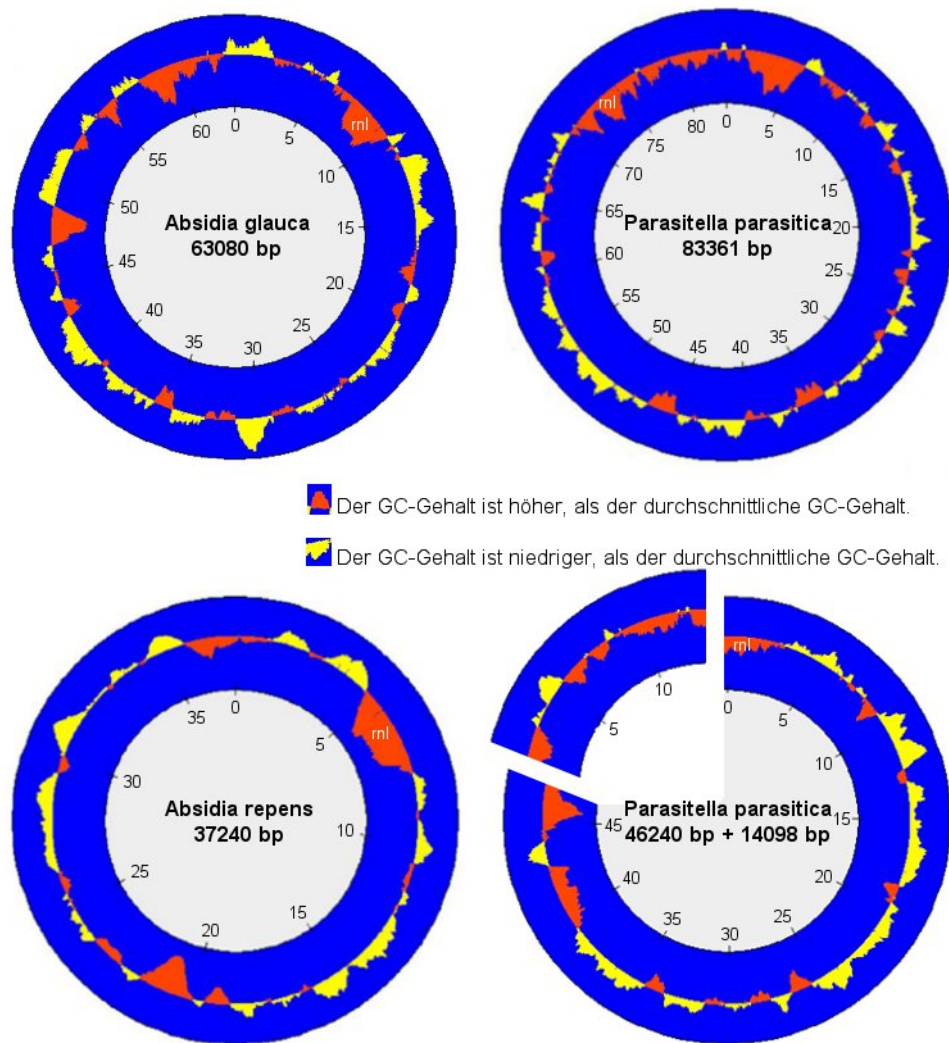


Abbildung 3.6: Ringförmige Darstellung des GC-Gehalts der mtDNAs zweier *Absidia*- und *Parasitella*-Stämme mit GC-Draw. Der GC-Gehalt wurde für eine Fenstergröße von 1 000 nt berechnet. Die obere rechte Abbildung zeigt den GC-Gehalt der mtDNA aus *Parasitella parasitica* CBS 412.66. Die untere rechte Abbildung zeigt den GC-Gehalt der mtDNA aus *Parasitella parasitica* NRRL 2501. Das größere Fragment des Rings entspricht dem mitochondrialen Scaffold 1. Das kleinere Fragment entspricht dem mitochondrialen Scaffold 2 (siehe Abbildung 3.3). Die Anordnung der beiden Scaffolds ist beliebig gewählt. „rnl“ markiert das Gen der großen rRNA-Untereinheit zum besseren Vergleich der Synthenie zwischen den mtDNAs.

Genomeigenschaften	<i>Absidia glauca</i>	<i>Parasitella parasitica</i>
Mittlere Contiglänge (bp)	3 866	1 792
Genomgröße (Mbp)	49	45
GC-Gehalt (%) gesamt	45	39
GC-Gehalt (%) in Exons	48	44
GC-Gehalt (%) in Introns	29	24
Anzahl an Genen	15 594	14 507
Mittlere Genlänge (bp)	1 275	1 226
Anzahl an Exons	69 292	51 862
Mittlere Exonlänge (bp)	286	342
Anzahl an Introns	53 831	37 546
Mittlere Anzahl an Introns pro Gen	4,5	3,6
Mittlere Intronlänge (bp)	88	82
Anzahl an tRNAs	232	161
Anzahl an repetitiven Einheiten	62 775	48 959
mitochondriale Genome		
Größe der mtDNA (kbp)	63	83
GC-Gehalt der mtDNA (%)	28	30
Anzahl an tRNAs in der mtDNA	24	26
Anzahl an Endonukleasen in der mtDNA	12	27

Tabelle 3.5: Übersicht der Eigenschaften der sequenzierten Genome von *Absidia glauca* und *Parasitella parasitica* CBS 412.66.

Tabelle 3.5 zeigt die wichtigsten Eigenschaften der beiden sequenzierten Genome von *Absidia glauca* und *Parasitella parasitica* CBS 412.66 im Vergleich.

3.1.5 Transponierbare Elemente im Kerngenom

Neben einfachen Wiederholungen, die oft nur aus einer längeren Abfolge aus ein oder zwei Nukleotiden bestehen, finden sich in den Genomen von *Absidia glauca* und *Parasitella parasitica* viele transponierbare Elemente. Ein großer Teil der repetitiven Sequenzen lässt sich keinem bekannten Typ zuordnen. Bei *Absidia glauca* sind 35 % von unbekanntem Typ, bei *Parasitella parasitica* sind es sogar 57 %.

Das Genom von *Absidia glauca* enthält fast 63 000 repetitive Sequenzen aus 16 Familien transponierbarer Elemente. Insgesamt machen diese repetitiven Sequenzen rund 15,6 Mbp des Genoms aus. Die meisten Repeats (9 679) sind DNA-Transposons (DNA/) und stammen aus der DNA/Crypton-Familie. An die 4,3 Mbp sind einfache Wiederholungen (26 839). Weitere häufig auftretende transponierbare Elemente stammen aus

den Familien LTR (long terminal repeat)/Gypsy (557), DNA/MuLE-MuDR (397) oder DNA/CMC-EnSpm (374). Die Zahlen in Klammern entsprechen der Anzahl gefundener Elemente des entsprechenden Typs.

Das Genom von *Parasitella parasitica* enthält fast 49 000 repetitive Elemente aus 20 Familien transponierbarer Elemente. Insgesamt machen diese repetitiven Sequenzen rund 10 Mbp des Genoms aus. Die meisten Repeats (1 904) sind Retrotransposons und stammen aus der LTR/Gypsy-Familie. An die 0,4 Mbp sind einfache Wiederholungen (1 552). Weitere häufig auftretende transponierbare Elemente stammen aus den Familien RC/Helitron (1 215), LTR/DIRS (965) oder DNA/TcMar-ISRm11 (950). Die Zahlen in Klammern entsprechen der Anzahl gefundener Elemente des entsprechenden Typs.

In *Absidia glauca* und *Parasitella parasitica* finden sich transponierbare Elemente aus drei Ordnungen: DIRS (*Dictyostelium* intermediate repeat sequence), LINE (long interspersed nuclear element) und LTR-Retrotransposons (long terminal repeat). Abbildung 3.7 zeigt die Klassifikation der identifizierten transponierbaren Elemente. In *Absidia glauca* finden sich 57 transponierbare Elemente aus der Gruppe der endogenen Retroviren (ERV). *Parasitella parasitica* besitzt keine transponierbaren Elemente dieser Art.

Bei *Parasitella parasitica* finden sich 965 DIRS-ähnliche Elemente. *Absidia glauca* besitzt keine transponierbaren Elemente dieser Art. Bei *Absidia glauca* und *Parasitella parasitica* finden sich LINEs aus unterschiedlichen Familien. *Parasitella parasitica* besitzt Elemente aus den Familien LINE/Jockey und LINE/R1, die es in *Absidia glauca* nicht gibt. Elemente aus der Familie LINE/RTE-BovB kommen hingegen nur bei *Absidia glauca* vor.

Absidia glauca und *Parasitella parasitica* besitzen Transposons aus der RC/Helitron-Familie, die sich über einen „rolling-circle“-Mechanismus vervielfältigen. Insgesamt besitzt *Parasitella parasitica* DNA-Transposons aus neun Klasse II-Familien. Bei *Absidia glauca* sind es sechs unterschiedliche Klasse II-Familien. In *Absidia glauca* kommen 38 transponierbare Elemente vom Typ *Merlin* vor, den es in *Parasitella parasitica* nicht gibt. In *Parasitella parasitica* gibt es transponierbare Elemente aus den Familien DNA/Ginger, DNA/MuLE-F, DNA/TcMar-ISRm11 und DNA/TcMar-Tc1, die es in *Absidia glauca* nicht gibt.

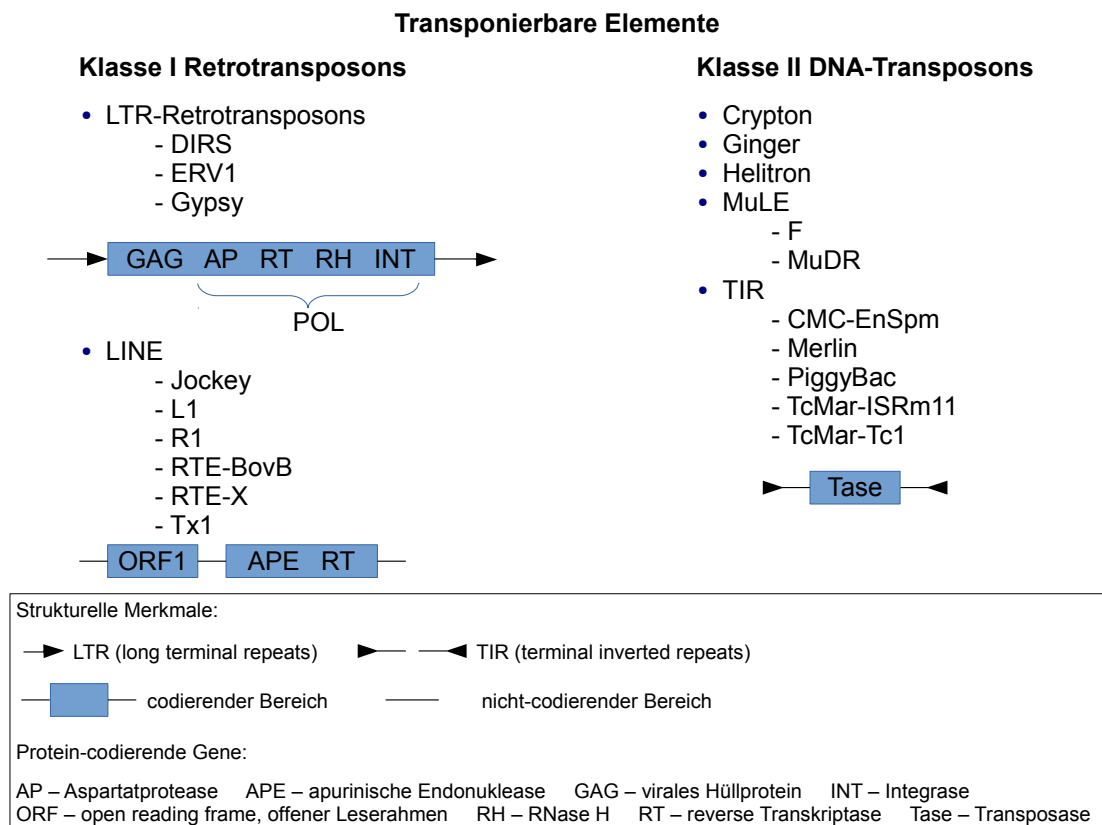


Abbildung 3.7: Klassifizierung transponierbarer Elemente in Eukaryoten.

In *Absidia glauca* und *Parasitella parasitica* finden sich Retrotransposons und DNA-Transposons aus verschiedenen Familien.

DIRS: *Dictyostelium* intermediate repeat sequence; LINE: long interspersed nuclear element; LTR: long terminal repeat; ERV: endogener Retrovirus.

Die Anordnung der strukturellen Merkmale und Protein-codierenden Domänen der drei als Beispiele ausgewählten Elemente *Gypsy*, *L1* und *Tc1-Mariner* stammt aus Wicker et al. (2007).

In *Absidia glauca* und *Parasitella parasitica* finden sich transponierbare Elemente aus Familien, für die horizontaler Gentransfer dokumentiert ist. Besonders in Metazoa, wie zum Beispiel *Drosophila*, wurden HGT-Ereignisse für LINE/*Jockey* (Mizrokhi und Mazo, 1990) und LINE/BovB (Župunski et al., 2001), sowie LTR/*Gypsy* (Terzian et al., 2000) Elemente gefunden. Tabelle 3.6 zeigt einen Vergleich der in den beiden Genomen identifizierten repetitiven Sequenzen. Für fettgedruckte Elemente ist horizontaler Gentransfer dokumentiert (Silva et al., 2004).

<i>Absidia glauca</i>			Repeatfamilien	<i>Parasitella parasitica</i>		
Gesamt-länge (bp)	Anzahl an Repeats	Anzahl an Repeattypen		Anzahl an Repeattypen	Anzahl an Repeats	Gesamt-länge (bp)
145 999	374	5	DNA/CMC-EnSpm	4	379	143 008
4 985 759	9 679	3	DNA/Crypton	8	510	209 338
			DNA/Ginger	2	76	28 581
9 405	38	1	DNA/Merlin			
			DNA/MuLE-F	1	80	33 074
135 658	397	6	DNA/MuLE-MuDR	7	677	211 695
29 626	71	1	DNA/PiggyBac	2	169	45 769
			DNA/TcMar-ISRm11	8	950	320 074
			DNA/TcMar-Tc1	5	384	105 439
			LINE	2	155	36 152
			LINE/Jockey	1	30	7 562
30 962	86	2	LINE/L1	9	681	248 314
60 678	172	2	LINE/L1-Tx1	6	655	234 368
			LINE/R1	1	97	45 596
16 690	45	1	LINE/RTE-BovB			
38 961	83	1	LINE/RTE-X	2	200	62 618
8 287	26	1	LTR			
			LTR/DIRS	8	965	290 461
20 100	57	1	LTR/ERV1			
215 806	557	4	LTR/Gypsy	22	1 904	693 135
62 724	141	1	RC/Helitron	12	1 215	386 962
106 637	2 324	3	Geringe Komplexität	3	1 419	64 168
4 265 620	26 839	1 995	Einfache Wiederholungen	1 552	10 515	380 634
5 442 089	21 886	142	Unbekannter Typ	267	27 898	6 552 157
Repetitive Elemente: 62 775				Repetitive Elemente: 48 959		
Repeattypen: 2 169				Repeattypen: 1 922		
Repeatfamilien: 16				Repeatfamilien: 20		
Gesamtlänge (bp): 15 575 001				Gesamtlänge (bp): 10 099 105		
Anteil am Genom: 32 %				Anteil am Genom: 22 %		

Tabelle 3.6: Übersicht der repetitiven Sequenzen in den sequenzierten Genomen von *Absidia glauca* und *Parasitella parasitica*. Aufbau und Eigenschaften der unterschiedlichen Familien werden in Abbildung 3.7 dargestellt. Transponierbare Elemente, für die horizontaler Gentransfer dokumentiert ist (Silva et al., 2004), sind fettgedruckt.

3.2 Strukturmodellierung von TSP1, TSP2 und TSP3

TSP1, TSP2 und TSP3 sind Enzyme der Trisporosäure-Biosynthese in Zygomyceten (Abbildung 1.5). TSP1 und TSP2 nehmen Trisporoide als Substrat auf und wandeln diese in andere Trisporoide um. Obwohl ihre Substrate nahezu identisch sind, ist ihre Proteinstruktur grundlegend verschieden. TSP1 ist eine Aldo-Keto-Reduktase mit einer TIM-Barrel-Struktur. TSP2 ist eine kurzkettige Dehydrogenase mit einer Rossmann-Faltung. TSP3 spaltet β -Carotin im ersten Schritt der Trisporosäure-Produktion. Hierbei handelt es sich um ein siebenblättriges β -Propeller-Protein. Abbildung 3.8 zeigt die drei verschiedenen Proteinstrukturen.

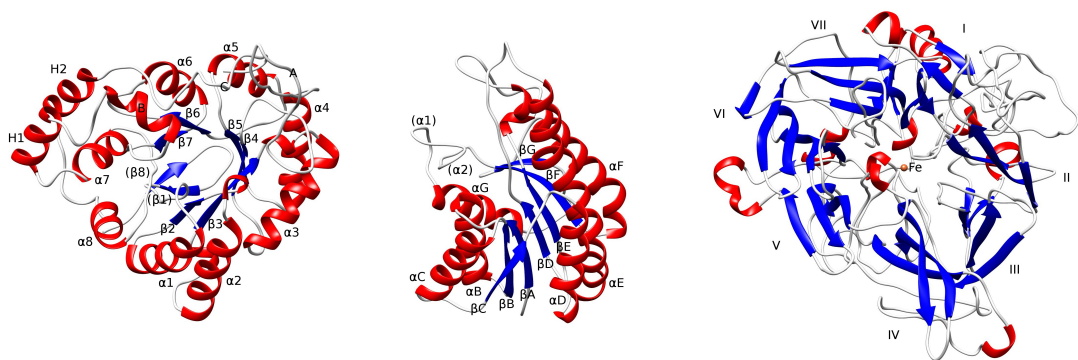


Abbildung 3.8: Vergleich der modellierten Proteinstrukturen für TSP1, TSP2 und TSP3 aus *Parasitella parasitica*. Rote Bereiche sind Helices. Blaue Bereiche sind Stränge. Die Nummerierung folgt der für Aldo-Keto-Reduktasen, kurzkettige Dehydrogenasen und Propeller-Proteine üblichen Schreibweise. Strukturelemente in Klammern fehlen in den meisten Protein-Modellen für Zygomyceten, treten aber bei Proteinen auf, die als Vorlage für die Homologie-Modellierung gedient haben.

3.2.1 4-Dihydromethyltrisporat-Dehydrogenase (TSP1)

Mit Hilfe einer BLAST-Suche mit der bereits identifizierten TSP1-Sequenz aus *Mucor mucedo* (UniProt: Q01213) wurden weitere TSP1-Sequenzen für alle Zygomyceten in der JGI MycoCosm-Datenbank (Stand 07.2016) identifiziert. Für *Absidia repens* NRRL 1336 (4), *Backusella circina* FSU 941 (9), *Blakeslea trispora* NRRL 2456 (2), *Hesseltinella vesiculosa* NRRL 3301 (1), *Lichtheimia hyalospora* FSU 10163 (7), *Mucor circinelloides* CBS 277.49 (6), *Mucor heterogamus* NRRL 1489 (2), *Mycotypha africana* NRRL 2978 (3), *Parasitella parasitica* NRRL 2501 (5), *Phycomyces blakesleeanus* NRRL 1555 (1), *Rhizopus microsporus* ATCC 52813 (2), *Rhizopus oryzae* RA 99-880 (1), *Syncephalastrum racemosum* NRRL 2496 (6), *Thamnidium elegans* NRRL 2467 (1) und *Umbelopsis ramanniana* AG (1) wurden insgesamt 51 TSP1-Sequenzen identifiziert. Hinzukommen weitere 28 TSP1-Sequenzen aus Zygomyceten, die in NCBI identifiziert wurden. Diese stammen aus *Lichtheimia corymbifera* FSU 9682 (7), *Lichtheimia ramosa* FSU 6197 (7), *Mucor ambiguus* NBRC 6742 (5), *Mucor circinelloides* 1006PhL (6) und *Rhizomu-*

cor pusillus NBRC 4578 (2). Auch in den von uns sequenzierten Zygomyceten *Absidia glauca* CBS 101.48 RVII-324 met⁻ (4) und *Parasitella parasitica* CBS 412.66 NGI-315 ade⁻ (6) wurden mehrere TSP1-Sequenzen identifiziert. In Klammern steht die Anzahl an TSP1-Sequenzen, die mindestens 95 % der Länge eines typischen TSP1-Proteins (321 Aminosäuren) aufweisen. In einigen Zygomyceten gibt es weitere kürzere Fragmente. Diese wurden für die Analysen nicht berücksichtigt. Die Anzahl an TSP1-Sequenzen einer Spezies variiert deutlich. Einige Zygomyceten, wie zum Beispiel *Hesseltinella vesiculosa* und *Phycomyces blakesleeanus*, besitzen nur ein TSP1. Andere Zygomyceten, wie zum Beispiel *Backusella circina*, besitzen überraschenderweise bis zu neun TSP1- oder TSP1-ähnliche Proteine. Abbildung 3.9 zeigt einen Neighbor-Joining-Baum, der mit Clustal Omega (Sievers et al., 2011) erstellt wurde. Alle 89 TSP1-Sequenzen aus Zygomyceten, sowie TSP1-ähnliche Sequenzen aus dem arbuskulären Mykorrhizapilz *Rhizophagus irregularis* und dem Ascomyceten *Ampelomyces quisqualis* sind enthalten. Es ist deutlich erkennbar, dass sich die verschiedenen TSP1-Sequenzen nicht nach dem Organismus gruppieren, aus dem sie stammen, sondern eher Protein-Untergruppen bilden. Die Sequenzidentität zwischen zwei TSP1-Isoformen einer Spezies ähnelt eher dem Verhältnis von TSP1-Sequenzen aus unterschiedlichen Pilzen, statt höhere Sequenzidentitäten aufzuweisen. Phylogenetische Analysen platzieren *Rhizomucor pusillus* eher in der Nähe von *Lichtheimia hyalospora* und *Phycomyces blakesleeanus* als von *Mucor circinelloides* (Hoffmann et al., 2013). Diese Klassifikation wird von den TSP1-Daten nicht bestätigt. Beide TSP1-Sequenzen aus *Rhizomucor pusillus* gruppieren sich mit TSP1-Sequenzen aus *Mucor circinelloides* und liegen in zwei getrennten Teilbäumen.

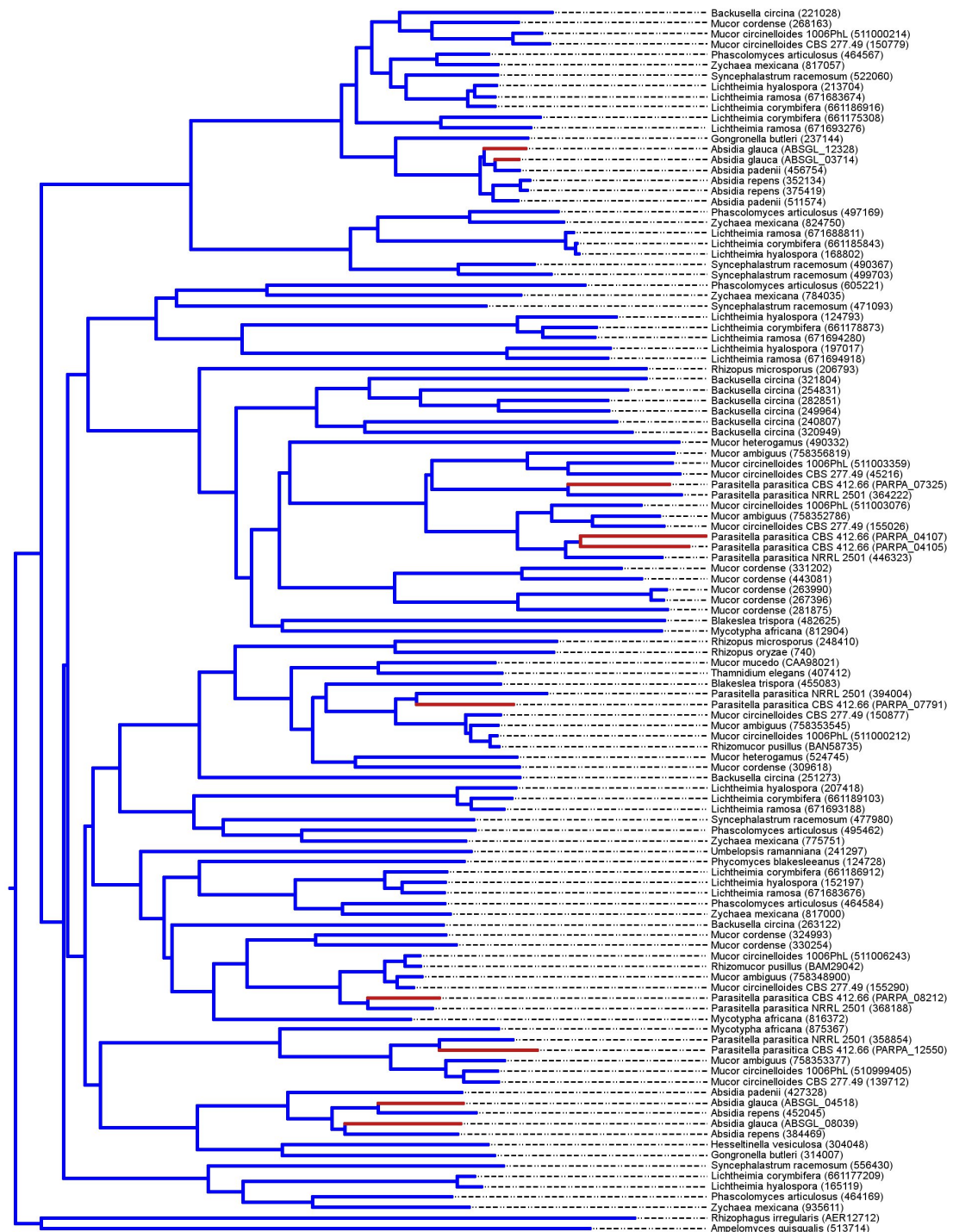


Abbildung 3.9: TSP1- und TSP1-ähnliche Proteine aus 28 verschiedenen Zygomyceten, dem arbuskulären Mykorrhizapilz *Rhizophagus irregularis* und dem Ascomyceten *Ampelomyces quisqualis*. Die Sequenzen aus *Absidia glauca* und *Parasitella parasitica* sind rot markiert.

TSP1-Sequenzen lassen sich in verschiedene Untergruppen einteilen. Diese TSP1-Untergruppen werden anhand der letzten acht Aminosäuren in der C-terminalen Region definiert. Die Sequenz-Logos in Abbildung 3.10 verdeutlichen die Unterschiede zwischen den drei Gruppen.

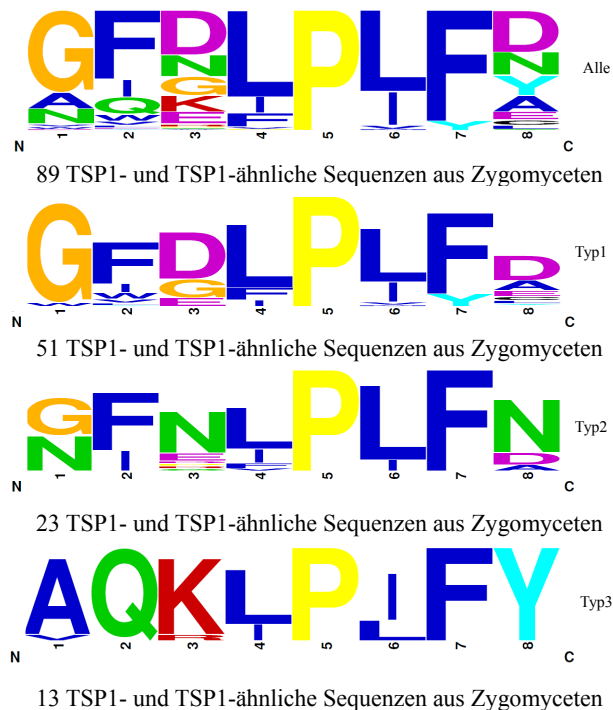


Abbildung 3.10: Sequenz-Logos der C-terminalen Sequenzmotive von TSP1-Proteinen. Das oberste Sequenzmotiv zeigt das Profil aller 89 TSP1-Proteine. Die unteren drei Sequenzmotive definieren drei TSP1-Untergruppen, die aus 51, 23 und 13 TSP1-Sequenzen bestehen.

Unter den besten BLAST-Treffern für die neu identifizierten TSP1-Proteine findet sich meist die TSP1-Sequenz aus *Mucor mucedo* (UniProt: Q01213). Bei *Lichtheimia hyalosporea* liegt die TSP1-Sequenz aus *Mucor mucedo* an zweiter Stelle hinter einer Xylose-Reduktase aus *Rhizomucor pusillus* (UniProt: I7HD84). Eine BLAST-Suche für diese Sequenz in UniProt ergab vier TSP1-Sequenzen innerhalb der besten fünf BLAST-Treffer und eine Xylose-Reduktase aus *Talaromyces emersonii* (UniProt: C5J3R6) auf Platz fünf. Dieses Ergebnis zeigt, dass es sich bei dieser Xylose-Reduktase aus *Rhizomucor pusillus* in Wirklichkeit um eine weitere TSP1-Sequenz handelt. Für *Rhizomucor pusillus* gibt es bereits eine als 4-Dihydromethyltrisorat-Dehydrogenase benannte Sequenz (UniProt: S0BD06). Ein Alignment der beiden Sequenzen liefert eine Sequenzidentität von 65 %. Ein Sequenzvergleich mit anderen TSP1-Sequenzen beweist, dass es sich sowohl bei der als TSP1 benannten Sequenz, als auch bei der als Xylose-Reduktase bezeichneten Sequenz, um TSP1-Proteine handelt. Die TSP1-Sequenz aus *Rhizomucor pusillus* ist am ähnlichsten zu einer TSP1-Sequenz aus *Mucor circinelloides* 1006PhL (JGI: 511000212) mit einer Sequenzidentität von 99 %.

Nr.	Protein-ID	Datenbank	Länge (AS)	MG (kDa)	α 5-Helix	C-terminales Ende	Aktives Zentrum
<i>Absidia glauca</i> (CBS 101.48 RVII-324 met-, (-)-Kreuzungstyp)							
Seq1	SAL98187	NCBI	321	36,92	AQL L MDMLTY	AQKLPIFY	D-45, Y-50, K-79, H-112
Seq2	SAL98947	NCBI	323	36,7	V M ALLDLFTY	GLDLPLFD	D-48, Y-53, K-82, H-115
Seq3	<u>SAM02260</u>	NCBI	321	36,83	V M ALLDLFTY	GFDLPLFD	D-46, Y-51, K-80, H-113
Seq4	SAM06439	NCBI	321	36,79	AQL L MDMLS Y	AQKLPIFY	D-45, Y-50, K-79, H-112
<i>Absidia repens</i> (NRRL 1336, (+)-Kreuzungstyp)							
Seq1	375419	JGI / MycoCosm	321	36,8	VQL L MDMLS Y	AQKLPIFY	D-45, Y-50, K-79, H-112
Seq2	452045	JGI / MycoCosm	321	36,67	V M ALLDLFTY	GLDLPLFD	D-46, Y-51, K-80, H-113
Seq3	384469	JGI / MycoCosm	321	36,6	V M ALLDLFTY	GFDLPLFD	D-46, Y-51, K-80, H-113
Seq4	352134	JGI / MycoCosm	322	36,87	VQL L MDMLS Y	AQKLPIFY	D-45, Y-50, K-79, H-112
<i>Backusella circina</i> (FSU 941, (+)-Kreuzungstyp)							
Seq1	251273	JGI / MycoCosm	321	36,44	VQSIL D MLTY	GFNLPLFD	D-46, Y-51, K-80, H-113
Seq2	263122	JGI / MycoCosm	323	36,68	VQLI L DLLTY	GFGLPLFC	D-47, Y-52, K-81, H-114
Seq3	249964	JGI / MycoCosm	323	36,86	VQSV L DLLTY	GFELPIFN	D-47, Y-52, K-81, H-114
Seq4	221028	JGI / MycoCosm	326	37,17	CQL L MDMLTY	VQKIPLFY	D-46, Y-51, K-80, H-113
Seq5	320949	JGI / MycoCosm	320	36,51	VQSTV D LLTY	DFGVPVFD	D-44, Y-49, K-78, H-111
Seq6	240807	JGI / MycoCosm	322	37,15	TQTIL D LLTY	VEVPPLFD	D-45, Y-50, K-79, H-112
Seq7	321804	JGI / MycoCosm	325	37,06	VQTIL D LLTY	GFELPMFD	D-47, Y-52, K-81, H-114
Seq8	254831	JGI / MycoCosm	323	36,68	TQSIL D LLTY	GFELPIFN	D-47, Y-52, K-81, H-114
Seq9	282851	JGI / MycoCosm	323	36,67	VQSV L DLLTY	GFELPIFD	D-47, Y-52, K-81, H-114
<i>Blakeslea trispora</i> (NRRL 2456, (+)-Kreuzungstyp)							
Seq1	<u>455083</u>	JGI / MycoCosm	321	36,92	CQA I MDLLTY	GFGIPLFD	D-46, Y-51, K-80, H-113
Seq2	482625	JGI / MycoCosm	324 *	37,18	VQLI I DLLAY	GVELPIFE	D-50, Y-55, K-84, H-117
<i>Hesseltinella vesiculosa</i> (NRRL 3301, (-)-Kreuzungstyp)							
Seq1	304048	JGI / MycoCosm	319	36,47	V M ML L DMFSY	GIDLPLFD	D-44, Y-49, K-78, H-111
<i>Lichtheimia corymbifera</i> (FSU 9682, (-)-Kreuzungstyp)							
Seq1	CDH51040	NCBI	321	36,1	VQLI L DLLTY	GFGLPLFC	D-45, Y-50, K-79, H-112
Seq2	CDH48981	NCBI	322	37,24	VQSM L DLFTY	GFDLPLFA	D-46, Y-51, K-80, H-113
Seq3	CDH51836	NCBI	320	36,04	AQ M LM D MLAY	GWDLPIFL	D-46, Y-51, K-80, H-111
Seq4	CDH58265	NCBI	322	37,78	VQLT L DLLTY	GIDIPLYE	D-47, Y-52, K-81, H-114
Seq5	CDH59815	NCBI	325	36,91	VQTIL D MLTY	GIDLPIFD	D-47, Y-52, K-81, H-114
Seq6	CDH51044	NCBI	322	36,78	VQL L MDMLTY	AQKLPIFY	D-46, Y-51, K-80, H-113
Seq7	CDH61395	NCBI	322	36,83	VQL L MDMLTY	AQRLPIFY	D-46, Y-51, K-80, H-113

Tabelle 3.7: Liste aller berücksichtigten TSP1-Sequenzen aus Zygomyceten.

Unterstrichene Sequenzen wurden experimentell als TSP1 identifiziert.

Fettgedruckte Aminosäuren unterscheiden die α 5-Helices in TSP1 und in der als Vorlage verwendeten Xylose-Reduktase aus *Candida tenuis* (Q/M statt A und A/S/T/V statt R).

MG: Molekulargewicht des Proteins

#: Intron-Exon-Grenzen wurden per Hand korrigiert.

*: Die Sequenz wurde am Anfang gekürzt.

N: Eine Aminosäure des aktiven Zentrums fehlt oder wurde gegen ein Arginin ausgetauscht.

Nr.	Protein-ID	Datenbank	Länge (AS)	MG (kDa)	α5-Helix	C-terminales Ende	Aktives Zentrum
<i>Lichtheimia hyalospora</i> (FSU 10163, (-)-Kreuzungstyp)							
Seq1	152197	JGI / MycoCosm	322	36,13	VQLILDLLTY	GFGLPLFC	D-46, Y-51, K-80, H-113
Seq2	207418	JGI / MycoCosm	322	37,08	VQSMLDLFTY	GFDLPLFA	D-46, Y-51, K-80, H-113
Seq3	168802	JGI / MycoCosm	320	36,03	AQMLMDMLAY	GWDLPIFL	D-46, Y-51, K-80, H-111
Seq4	197017	JGI / MycoCosm	320	37,02	VQLTLDLLTY	GIDIPLYE	D-45, Y-50, K-79, H-112
Seq5	124793	JGI / MycoCosm	322	37,73	VQLTLDLLTY	GVDLPVYD	D-47, Y-52, K-81, H-114
Seq6	165119	JGI / MycoCosm	325	36,96	VQTILDMLTY	GIDLPIFD	D-47, Y-52, K-81, H-114
Seq7	213704	JGI / MycoCosm	322	36,79	VQLLMDMLTY	AQKLPIFY	D-46, Y-51, K-80, H-113
<i>Lichtheimia ramosa</i> (FSU 6197, (-)-Kreuzungstyp)							
Seq1	CDS12710	NCBI	321	36,13	VQLILDLLTY	GFGLPLFC	D-45, Y-50, K-79, H-112
Seq2	CDS07919	NCBI	322	37,14	VQSMLDLFTY	GFDLPLFA	D-46, Y-51, K-80, H-113
Seq3	CDS03129	NCBI	320	36,05	AQMLMDMLAY	GWDLPIFL	D-46, Y-51, K-80, H-111
Seq4	CDS09639	NCBI	321	37,02	VQMTLDLLTY	GIDIPLYE	D-46, Y-51, K-80, H-113
Seq5	CDS09007	NCBI	322	37,67	VQLTLDLLTY	GIDLPLYD	D-47, Y-52, K-81, H-114
Seq6	CDS12708	NCBI	322	36,81	VQLLMDMLTY	AQKLPIFY	D-46, Y-51, K-80, H-113
Seq7	CDS08007	NCBI	322	36,89	VQLLMDMLTY	AQKLPIFY	D-46, Y-51, K-80, H-113
<i>Mucor ambiguus</i> (NBRC 6742, (+)-Kreuzungstyp)							
Seq1	GAN04332	NCBI	321	36,86	CQAILDMLTY	GFGLPLFD	D-46, Y-51, K-80, H-113
Seq2	GAN08850	NCBI	322	36,65	VQLILDLLTY	NFNLPFLN	D-47, Y-52, K-81, H-114
Seq3	GAN01222	NCBI	321	36,42	VQLTLDLLSY	GFDFPIYD	D-47, Y-52, K-81, H-114
Seq4	GAN05122	NCBI	321	36,39	VQSTLDLLSY	GFDFPLFA	D-47, Y-52, K-81, H-114
Seq5	GAN04526	NCBI	318	36,47	AQLIMDLLTY	NINIPFLN	D-45, Y-50, K-79, H-112
<i>Mucor circinelloides</i> (CBS 277.49, (-)-Kreuzungstyp)							
Seq1	150877	JGI / MycoCosm	321	36,95	CQAILDMLTY	GFGLPLYD	D-46, Y-51, K-80, H-113
Seq2	155290	JGI / MycoCosm	322	36,66	VQLILDLLTY	NFNLPFLD	D-47, Y-52, K-81, H-114
Seq3	155026	JGI / MycoCosm	321	36,49	VQSTLDLLSY	GFNFPLFA	D-47, Y-52, K-81, H-114
Seq4	45216	JGI / MycoCosm	321	36,46	VQSTLDLLSY	GFEFPLFN	D-47, Y-52, K-81, H-114
Seq5	150779	JGI / MycoCosm	325	37,39	CQLLMDMLTY	AQKIPLFY	D-46, Y-51, K-80, H-113
Seq6	139712	JGI / MycoCosm	318	36,45	AQLIMDLLTY	NIDIPLFN	D-45, Y-50, K-79, H-112
<i>Mucor circinelloides</i> (1006PhL, (+)-Kreuzungstyp)							
Seq1	EPB81768	NCBI	321	36,85	CQAILDMLTY	GFGLPLFD	D-46, Y-51, K-80, H-113
Seq2	EPB87560	NCBI	322	36,71	VQLILDLLTY	NFNLPFLN	D-47, Y-52, K-81, H-114
Seq3	EPB84490	NCBI	321	36,49	VQSTLDLLSY	GFDFPLFA	D-47, Y-52, K-81, H-114
Seq4	EPB84764	NCBI	321	36,45	VQLTLDLLSY	GFEFPLFD	D-47, Y-52, K-81, H-114
Seq5	EPB81770	NCBI	325	37,27	CQLLMDMLTY	AQKIPLFY	D-46, Y-51, K-80, H-113
Seq6	EPB81046	NCBI	318	36,46	AQLIMDLLTY	NINIPFLN	D-45, Y-50, K-79, H-112
<i>Mucor heterogamus</i> (NRRL 1489, homothallisch)							
Seq1	524745	JGI / MycoCosm	321 #	37,1	CQAILDMLTY	GFNLPLFD	D-46, Y-51, K-80, H-113
Seq2	490332	JGI / MycoCosm	320 #	36,86	VQSIVDLLTY	GFDFPLFE	D-48, Y-53, K-82, H-115

Tabelle 3.7: (Fortsetzung)

Nr.	Protein-ID	Datenbank	Länge (AS)	MG (kDa)	α 5-Helix	C-terminales Ende	Aktives Zentrum
<i>Mucor mucedo</i> (CBS 109.16, (-)-Kreuzungstyp)							
Seq1	CAA98021	NCBI	321	36,49	CQAILDLLTY	GFGLPLFD	D-46, Y-51, K-80, H-113
<i>Mycotypha africana</i> (NRRL 2978, homothallisch)							
Seq1	816372	JGI / MycoCosm	322	37,05	VQMILDLLTY	GFGLPLFD	D-47, Y-52, K-81, H-114
Seq2	875367	JGI / MycoCosm	319	36,7	CQIIMDLLTY	GIPVPLFN	D-46, Y-51, K-80, H-113
Seq3	812904	JGI / MycoCosm	322	37,38	CQLILDLLTY	GFELPIFE	D-48, Y-53, K-82, H-115
<i>Parasitella parasitica</i> (CBS 412.66 NGI-315 ade-, (+)-Kreuzungstyp)							
Seq1	CEP13671	NCBI	321	36,96	CQAILDLLSY	GFGLPLFD	D-46, Y-51, K-80, H-113
Seq2	CEP14056	NCBI	322	36,60	VQLILDLLTY	NFNLPLFN	D-47, Y-52, K-81, H-114
Seq3	CEP10423	NCBI	321	36,54	VQSTLDLLTY	GFDFPVFA	D-47, Y-52, K-81, H-114
Seq4	CEP10425	NCBI	322	36,63	VQSTLDLLTY	GFDFPLFA	D-47, Y-52, K-82, H-115
Seq5	CEP13270	NCBI	321	36,36	VQSILDLLTY	GFNLPLFN	D-47, Y-52, K-81, H-114
Seq6	CEP18248	NCBI	322	37,05	AQLIMDLLTY	NIRIPLFN	D-45, Y-50, <u>N-79</u> , H-116
<i>Parasitella parasitica</i> (NRRL 2501, (-)-Kreuzungstyp)							
Seq1	394004	JGI / MycoCosm	328 *	37,61	CQAILDLLSY	GFGLPLFD	D-53, Y-58, K-87, H-120
Seq2	368188	JGI / MycoCosm	322	36,73	VQLILDLLTY	NFNLPLFN	D-47, Y-52, K-81, H-114
Seq3	446323	JGI / MycoCosm	322	36,44	VQSTLDLLSY	GFDLPLFA	D-47, Y-52, K-81, H-114
Seq4	364222	JGI / MycoCosm	321	36,58	VQSTLDLLSY	GFNLPLFN	D-47, Y-52, K-81, H-114
Seq5	358854	JGI / MycoCosm	327 #	37,25	AQLIMDLLTY	NISIPLFN	D-31, Y-36, K-65, H-98
<i>Phycomyces blakesleeana</i> (NRRL 1555, (+)-Kreuzungstyp)							
Seq1	124728	JGI / MycoCosm	323	36,52	VQLILDLLTY	GFNLPLFA	D-47, Y-52, K-81, H-114
<i>Rhizomucor pusillus</i> (NBRC 4578, (+)-Kreuzungstyp)							
Seq1	BAN58735	NCBI	321	36,92	CQAILDMLTY	GFGLPLFD	D-46, Y-51, K-80, H-113
Seq2	BAM29042	NCBI	322	36,72	VQLILDLLTY	NFNLPLFN	D-47, Y-52, K-81, H-114
<i>Rhizopus microsporus</i> (ATCC 52813, (+)-Kreuzungstyp)							
Seq1	248410	JGI / MycoCosm	321	36,97	CQALIDLLEY	WFDLPLFA	D-46, Y-51, K-80, H-113
Seq2	206793	JGI / MycoCosm	322	37,37	VQSILDILTY	GYDFPIFD	D-48, Y-53, K-82, H-115
<i>Rhizopus oryzae</i> (RA 99-880, (+)-Kreuzungstyp)							
Seq1	740	JGI / MycoCosm	320	36,5	TQALVDMLTY	WFDLPLFA	D-46, Y-51, K-80, H-113
<i>Syncephalastrum racemosum</i> (NRRL 2496, homothallisch)							
Seq1	471093	JGI / MycoCosm	323	37,01	VQLILDVLTY	GFELPLFD	D-46, Y-51, K-80, H-113
Seq2	477980	JGI / MycoCosm	321	37,01	VQSIIELLSY	GFDLPLFS	D-46, Y-51, K-80, H-113
Seq3	490367	JGI / MycoCosm	322	36,15	VQMLMDLLSY	GWDLPIFY	D-46, Y-51, K-80, H-113
Seq4	499703	JGI / MycoCosm	322	36,22	VQSLMDLLSY	GWDLPIFY	D-46, Y-51, K-80, H-113
Seq5	556430	JGI / MycoCosm	326	37,15	VQTLLEDLVY	GVDLPIFD	D-47, Y-52, K-81, H-114
Seq6	522060	JGI / MycoCosm	322	36,69	VQLLMDMLTY	AQKLPLFY	D-46, Y-51, K-80, H-113
<i>Thamnidium elegans</i> (NRRL 2467, (+)-Kreuzungstyp)							
Seq1	407412	JGI / MycoCosm	321	36,72	CQAILDLLTY	GFNLPLFD	D-46, Y-51, K-80, H-113
<i>Umbelopsis ramanniana</i> (AG, (-)-Kreuzungstyp)							
Seq1	241297	JGI / MycoCosm	321	36,07	VQLLMDLLTY	GFNLPIFN	D-45, Y-50, K-79, H-112

Tabelle 3.7: (Fortsetzung)

Die Xylose-Reduktase ist am ähnlichsten zu einer TSP1-Sequenz aus *Mucor circinelloides* 1006PhL (JGI: 511006243) mit einer Sequenzidentität von 98 %.

Alle TSP1-Proteine beinhalten dieselben Proteindomänen, wie das erste modellierte TSP1-Protein aus dem Modell-Zygomyceten *Mucor mucedo* (Ellenberger et al., 2013). TSP1 ist eine Aldo-Keto-Reduktase (Czempinski et al., 1996) und beinhaltet die für diese Enzyme typischen Proteindomänen (InterPro: IPR001395, IPR018170, IRP020471, IPR023210). Tabelle 3.7 enthält alle identifizierten vollständigen TSP1-ähnlichen Sequenzen. Sie alle haben eine Länge von rund 320 Aminosäuren und ein Molekulargewicht (MG) von circa 37 kDa. Neben den entscheidenden Regionen der α 5-Helix und des C-terminalen Endes wird auch die Lage des aktiven Zentrums der Proteine aufgeführt. Die Zugänglichkeit des aktiven Zentrums variiert deutlich zwischen den einzelnen TSP1-Isomeren. Ein Vergleich der identifizierten TSP1-Proteine in Gruppe 1 ergab Sequenzidentitäten zwischen 56 % und 97 %. Im Durchschnitt lag die Sequenzidentität bei rund 67 %. Abbildung 3.11 zeigt ein Sequenzalignment einiger TSP1-Sequenzen aus Gruppe 1.



Abbildung 3.11: Sequenzalignment der TSP1-Proteine aus Gruppe 1 einiger Zygomyceten.

Rot markierte Aminosäuren bilden das aktive Zentrum der Enzyme. Blau markierte Aminosäuren sind an der Bindung des Cofaktors NADP beteiligt. Orange markierte Aminosäuren sind an der Bindung des Trisporoid-Liganden beteiligt. Kästchen markieren für die Dimerisierung wichtige Bereiche, eine flexible Schleifen-Region, die α 5-Helix und das C-terminale Ende.

	flexible Schleife	α5-Helix
Rhizopus microsporus_248410	YVDPSEKYPPEWFRKGR-DTIELEAS	MHECWAEMEKLVDGLVRNIGVGNFCQALIDL
Rhizopus oryzae_740	YVDPSQKYPPEWFKGNS-TAIEIESS	MHECWAEMERLVNDGLSRNIGVGNFTQALVDM
Backusella circina_251273	FVDFDTKYPPPEWYVPDA-TTLTYERS	THETWAEEMEKLVDAGLVRNIGVANFNVSQILDM
Mucor mucedo_CAA98021	YVDPATVYPPGWYVGDA-KSLQFEQS	THECWAELEKIVDAGLARNIGVANFNQAILDL
Thamnidium elegans_407412	YVSPADNYPGWYVGDA-KSIAFERS	THECWAELEKIVDAKLTRNIGVANFNQAILDL
Blakeslea trispora_455083	YVDPSVNYPPEWYVPGA-KTLTLERS	THECWAEEMEKLVEANLTRNIGVANFNQAIMDL
Parasitella parasitica (CBS 412.66)_CEP13671	YVDPAVSYPPEWYVPHA-DKLNFRS	THECWAEEMEKLVAKLTRNIGVANFNQAILDL
Mucor circinelloides (CBS 277.49)_150877	YVDPSVNYPPEWYVPGA-DKLTYSRS	THECWAEEMEKLVEAKLTRNIGVANFNQAILDM
Rhizomucor pusillus_BAN58735	YVDPSVNYPPEWYVPGA-EKLTYSRS	THECWAEEMEKLVEAKLTRNIGVANFNQAILDM
Absidia glauca_SAM02260	YVDPAVKYPGGWLQDGO-TSIAYSRS	THVCWAEEMEKLVDKVARNIGVSNFNMALLDL
Hesseltinella vesiculosa_304048	HVDPKVKYPGGWENPET-KKMEYQRS	LHLCWAEEMEKLVDAGLVRNIGVSNFNMMLDM
Umbelopsis ramanniana_241297	HVAIETAYPPGWFPGE-EKVTLDRA	MHECWAEEKLVEDGLARNIGVSNFNVQLLMDL
Syncephalastrum racemosum_477980	FVPFDKAYPPGWYQPGG-NKLEYERS	THQTWPEMEKLVDGLVRNIGVSNFNVQSIIEI
Lichtheimia hyalospora_152197	HVALEAAYPPGWYQPTGATEIEFERS	MQDCWREMEKIAESGKAKNIGVSNFNVQLIIDL
Phycomyces blakesleeanus_124728	NVEKSAAYPPGWYAPNGATAIEFERS	MHECWRELEKIAESGLAKNIGVSNFNVQLIIDL
	* ** * : : *	: * * : : : * : * : * *
Rhizopus microsporus_248410	LYAKIKPAVLQIELHPYLPQENLCKWVKAQGIHITA	SSSGPASYVTGEH-GKQAPSL
Rhizopus oryzae_740	LYAKIKPAVLQIELHPYLPQAEITKWVKSQGIHITA	SSSGPASYVTGEH-GKRAAPL
Backusella circina_251273	LYCKKYPVSVLQVEIHYPYLPQERLVKVVQAQGIQITA	SSSGPASYVTGEH-GKQAPSL
Mucor mucedo_CAA98021	LYARIKPAVLQIELHPYLPQERLVKVVKEQGIQITA	SSSGPASYVTGEH-GKTYTSL
Thamnidium elegans_407412	LYARIKPAVLQIELHPYLPQERLVKVVREQDIHITA	SSSGPASYVTGEH-GKTYTSL
Blakeslea trispora_455083	LYAKIKPSVLQIELHPYLPQERLVKVVQDQGIQVTA	SSSGPASYVTGEH-GKDYQSL
Parasitella parasitica (CBS 412.66)_CEP13671	LSYSKIKPAVLQIEIHPLLPQERLVKVVQEQGIQVTA	SSSGPASYVTGEH-GKTYTSL
Mucor circinelloides (CBS 277.49)_150877	LYAKIKPAVLQIEIHPLLPQERLVKVVQEQGIQVTA	SSSGPASYVTGEH-GKTYQSL
Rhizomucor pusillus_BAN58735	LYAKIKPAVLQIEIHPLLPQERLVKVVQEQGIQVTA	SSSGPASYVTGEH-GKTYQSL
Absidia glauca_SAM02260	FTYCRYKPAVLQVELHPYLPQKRLVWVQKQGVAVTA	SSSGPASYVTGEH-GKTAKPL
Hesseltinella vesiculosa_304048	FSYCRYKPAVLQVELHPYLPQKRLVWVQKQGVAVTA	SSSGPASYVTGEH-GKTAKPL
Umbelopsis ramanniana_241297	LYAKIKPAVLQVELHPYLPQKRLVWVQKQGVAVTA	SSSGPASYVTGEH-GKTAKPL
Syncephalastrum racemosum_477980	LSYCKYKPAVLQVENHPYLPQKRLVWVQKQGVAVTA	SSSGPASYVTGEH-GKTAKPL
Lichtheimia hyalospora_152197	LYANIKPAVLQVELHPYLPQKRLVWVQKQGVAVTA	SSSGPASYVTGEH-GKTAKPL
Phycomyces blakesleeanus_124728	LYANIKPAVLQVELHPYLPQKRLVWVQKQGVAVTA	SSSGPASYVTGEH-GKTAKPL
	: * : * : * : * : * : * : * : * : *	: * : * : * : * : * : * : * : * : *
Rhizopus microsporus_248410	LDHDIIVNSVANKHKVSAGQVLLRWALDREYVVI	KSQNEQRMKDNFNILDIKLDNEDNKA
Rhizopus oryzae_740	LEHDVKSADKHKVSAGQVLLRWALDREYVVI	KSQNEQRMKDNFNILDIKLDNEDNKA
Backusella circina_251273	LDHADVSSIASKHGVTGQVLLRWALDREYVVI	KSQNEQRMKDNFNILDIKLDNEDNKA
Mucor mucedo_CAA98021	LEHADVKSADKHKVSAGQVLLRWALDREYVVI	KSQNEQRMKDNFNILDIKLDNEDNKA
Thamnidium elegans_407412	LEHADVKSADKHKVSAGQVLLRWALDREYVVI	KSQNEQRMKDNFNILDIKLDNEDNKA
Blakeslea trispora_455083	LEHDQTKSISGNKHNVSAGQVLLRWALDREYVVI	KSQNEQRMKDNFNILDIKLDNEDNKA
Parasitella parasitica (CBS 412.66)_CEP13671	LANADVKSADKHKVSAGQVLLRWALDREYVVI	KSQNEQRMKDNFNILDIKLDNEDNKA
Mucor circinelloides (CBS 277.49)_150877	LANDNVKSADKHNITTQVLLRYALDRNIAVIV	KSQNEQRMKDNFNILDIKLDNEDNKA
Rhizomucor pusillus_BAN58735	LANDNVKSADKHNITTQVLLRYALDRNIAVIV	KSQNEQRMKDNFNILDIKLDNEDNKA
Absidia glauca_SAM02260	LSHELVDIATKYEKSTAQVLLRWVQKQGVAVTA	KSQNEQRMKDNFNILDIKLDNEDNKA
Hesseltinella vesiculosa_304048	LTHTETIQVAKKYSKSPAQVLLRWVQKQGVAVTA	KSQNEQRMKDNFNILDIKLDNEDNKA
Umbelopsis ramanniana_241297	LEHDVTSNIAKAKHNVSAGQVLLRWVQKQGVAVTA	KSQNEQRMKDNFNILDIKLDNEDNKA
Syncephalastrum racemosum_477980	LEHENVESIAKKGKTTGQVLLRWVQKQGVAVTA	KSQNEQRMKDNFNILDIKLDNEDNKA
Lichtheimia hyalospora_152197	LGHKVITDIASKHGKTAQVLLRWVQKQGVAVTA	KSQNEQRMKDNFNILDIKLDNEDNKA
Phycomyces blakesleeanus_124728	LEHEVKSIAKKGKTAGQVLLRWVQKQGVAVTA	KSQNEQRMKDNFNILDIKLDNEDNKA
	* . : . : . : * : * : * : * : * : *	* . : . : . : * : * : * : * : * : *
Rhizopus microsporus_248410	LDLSKFSNLRFDPKVM-FDLPLFA	
Rhizopus oryzae_740	LDLKS-NQRFNDPLVM-FDLPLFA	
Backusella circina_251273	LGALSS-NQRFNDPMHF-GFNLPLFD	
Mucor mucedo_CAA98021	LDLSKT-NQRFNDPMTF-GFGLPLFD	
Thamnidium elegans_407412	LDLSKT-NQRFNDPMTF-GFNLPLFD	
Blakeslea trispora_455083	LDALRT-NQRFNDPMIF-GFGLPLFD	
Parasitella parasitica (CBS 412.66)_CEP13671	LNSLKT-NQRFNDPMIF-GFGLPLFD	
Mucor circinelloides (CBS 277.49)_150877	LNNLKT-NQRFNDPMIF-GFGLPLFD	
Rhizomucor pusillus_BAN58735	LNNLKT-NQRFNDPMIF-GFGLPLFD	
Absidia glauca_SAM02260	LNNLKT-NQRFNDPMIF-GFGLPLFD	
Hesseltinella vesiculosa_304048	LNNLKT-NQRFNDPMIF-GFGLPLFD	
Umbelopsis ramanniana_241297	LNNLKT-NQRFNDPMIF-GFGLPLFD	
Syncephalastrum racemosum_477980	LNNLKT-NQRFNDPMIF-GFGLPLFD	
Lichtheimia hyalospora_152197	LNNLKT-NQRFNDPMIF-GFGLPLFD	
Phycomyces blakesleeanus_124728	LNNLKT-NQRFNDPMIF-GFGLPLFD	
	: : ***: : : : : *	

C-terminales Ende

Abbildung 3.11: (Fortsetzung)

Die meisten Aminosäuren, die für die Cofaktor-Bindung verantwortlich sind, sind in den TSP1-Sequenzen konserviert. Das aktive Zentrum besteht, mit nur einer Ausnahme, aus einem Aspartat, einem Tyrosin, einem Lysin und einem Histidin. Die Lage und der Abstand zwischen diesen Aminosäuren variieren leicht. Die Aminosäuren, die für die Bindung eines Trisporoid-Liganden notwendig sind, unterscheiden sich an einigen Positionen (Tabelle 3.7). Diese Unterschiede in den TSP1-Sequenzen ermöglichen es, Trisporoide aus verschiedenen Serien zu binden (siehe Abschnitt 1.1.2). Eine Überlagerung der Proteinmodelle (Abbildung 3.12) und ein multiples Struktur-Alignment der TSP1-Proteine (Abbildung 3.13) zeigen die deutliche Übereinstimmung der Grundstruktur der TSP1-Proteine. In Abbildung 3.12 (B) sind die Bereiche, die eine größere Variabilität aufweisen, im Detail dargestellt.

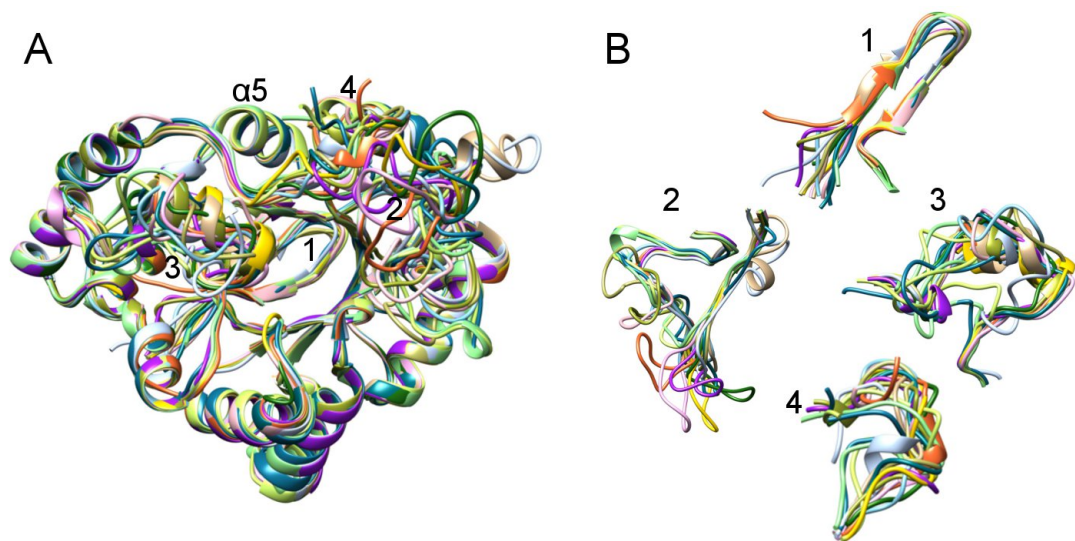
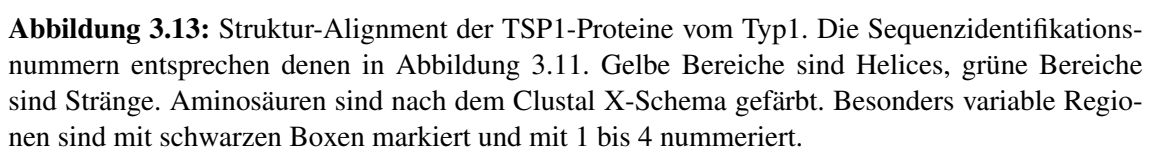


Abbildung 3.12: (A) Überlagerung der Proteinstrukturen von TSP1-Sequenzen vom Typ1 aus *Absidia glauca*, *Backusella circina*, *Blakeslea trispora*, *Hesseltinella vesiculosa*, *Lichtheimia hyalospora*, *Mucor circinelloides* (CBS 277.49), *Mucor mucedo*, *Parasitella parasitica* (CBS 412.66), *Phycomyces blakesleeanus*, *Rhizomucor pusillus*, *Rhizopus microsporus*, *Rhizopus oryzae*, *Syncephalastrum racemosum*, *Thamnidium elegans* und *Umbelopsis ramanniana*. Die Sequenzidentifikationsnummern entsprechen denen in Abbildung 3.11. $\alpha 5$ markiert die $\alpha 5$ -Helix, die bei der Dimerisierung der Xylose-Reduktase in *Candida tenuis* eine wichtige Rolle spielt. (B) Die Ziffern 1-4 bezeichnen variable Regionen der Strukturen. Diese Bereiche werden hier im Detail dargestellt. Die Farben entsprechen denen in Abbildung 3.13.

Region 1 besteht aus den ersten 18 Aminosäuren des Alignments. Region 2 und 3 befinden sich in der Mitte der Proteine. Region 2 beschreibt eine Schleife, die für die Substratspezifität und Enzymaktivität von TSP1 entscheidend ist. Diese Region bildet zusammen mit der Region 3, die sich auf der gegenüberliegenden Seite des Proteins befindet, die Substrat-Bindetasche aus. Region 4 besteht aus den letzten 14 Aminosäuren des Alignments. Eine zweite kurze Helix in diesem Bereich unterscheidet zwei verschiedene TSP1-Varianten in *Backusella circina*, *Mucor circinelloides* und *Rhizomucor pusillus* (Abbildung 3.15).



Interessanterweise liegen nur die Gene für TSP1 von *Typ1* in einem ähnlichen genomischen Umfeld, wie das *tsp1* aus *Parasitella parasitica*. Hier liegt unterhalb des *tsp1* ein Gen, das für eine Acyl-CoA-Thioesterase codiert. Dieses Gen wird in der entgegengesetzten Richtung transkribiert. Schultze et al. (2005) beschreiben einen Antisense-Mechanismus, der an dieser Stelle für eine posttranskriptionale Regulation des *tsp1* verantwortlich ist. Das Transkriptionslevel von *tsp1* und der benachbarten Acyl-CoA-Thioesterase hängen stark von parasitischen und sexuellen Interaktionen ab.

Die TSP1-Isoformen CEP10423 und CEP10425 aus *Parasitella parasitica* liegen auf demselben Scaffold fast direkt hintereinander. Ihre Proteinsequenzen ergeben eine Sequenzidentität von 88 %. PARPA_00452 ist ein Genfragment aus *Parasitella parasitica* auf einem kurzen Contig mit einer Sequenzidentität von 100 % zu CEP10423. Auch ein zweites Genfragment, CEP20050 aus *Parasitella parasitica*, zeigt die höchste Sequenzidentität zu CEP10423 (83 %). Andere *tsp1*-ähnliche Gene liegen auf anderen Scaffolds mit abweichendem genomischem Umfeld. Ihre Sequenzidentität ist höher zu TSP1-Sequenzen aus anderen Zygomyceten (bis zu 91 %), als zu anderen TSP1 aus demselben Organismus (im Durchschnitt 53 %).

Strukturalignments von TSP1- und TSP1-ähnlichen Proteinen ergaben RMSD-Werte zwischen 0.019 Å und 0.236 Å. Ein Struktur-Alignment von TSP1-Proteinen und der Proteinvorlage der Modellierung, der Xylose-Reduktase XYL1 (PDB: 1K8C) ergab RMSD-Werte zwischen 0.319 Å und 0.355 Å. Der Q-Wert der Struktur-Alignments unterstützt die Annahme, dass es sich bei den identifizierten Sequenzen wirklich um TSP1-Proteine handelt. Ein Q-Wert von eins, wie er für das Struktur-Alignment der TSP1-Proteine aus *Mucor circinelloides* und *Rhizomucor pusillus* zustande kommt, beschreibt zwei identische Proteinstrukturen. Die Q-Werte sind für alle TSP1-Alignments größer als 0.943 und bestätigen die strukturelle Ähnlichkeit der Enzyme.

Zygomyceten mit einem TSP1-Protein

Hesseltinella vesiculosa, *Mucor mucedo*, *Phycomyces blakesleeana*, *Rhizopus oryzae*, *Thamnidium elegans* und *Umbelopsis ramanniana* besitzen nur ein TSP1-Protein. Für *Mucor mucedo* wurde das komplette Genom noch nicht sequenziert. Die Verwandtschaft zu anderen Zygomyceten wie *Mucor circinelloides* lässt vermuten, dass es weitere TSP1-Proteine in diesem Pilz gibt. Abbildung 3.14 zeigt die modellierten Proteinstrukturen und die Ergebnisse aus dem Protein-Ligand-Docking für diese TSP1-Varianten.

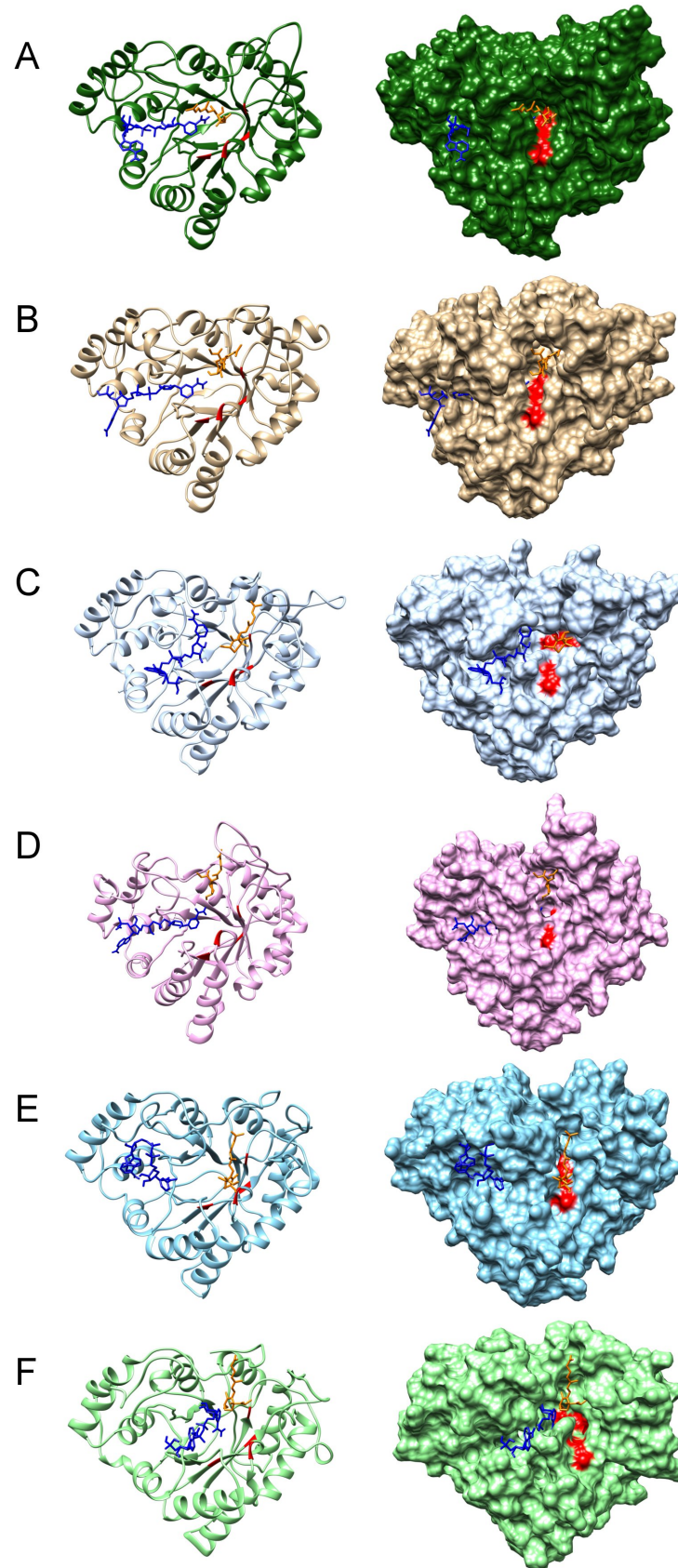


Abbildung 3.14: Modellierte TSP1-Strukturen für (A) *Hesseltinella vesiculosa*, (B) *Mucor mucedo*, (C) *Phycomyces blakesleeanus*, (D) *Rhizopus oryzae*, (E) *Thamnidium elegans* und (F) *Umbelopsis ramanniana*. Alle Proteinstrukturen beinhalten das beste Modell des Protein-Ligand-Dockings mit NADP (blau) und einem Trisporoid (orange). Das aktive Zentrum der Enzyme ist rot markiert.

Die Konformation des Trisporoid-Liganden innerhalb der Bindetasche unterscheidet sich zwischen den verschiedenen Pilzen. Die meisten Bindetaschen sind sehr eng. Bei *Thamnidium elegans* ist sie sogar für das Trisporoid geschlossen. Bei *Umbelopsis ramaniana* hingegen sind die Bindetaschen für NADP und Trisporoid miteinander verbunden und liegen offen zugänglich an der Oberfläche des Proteins.

Zygomyceten mit bis zu neun TSP1-Proteinen

Absidia glauca, *Absidia repens*, *Blakeslea trispora*, *Rhizomucor pusillus* und *Rhizopus microsporus* besitzen zwischen zwei und vier TSP1-Sequenzen. *Lichtheimia corymbifera*, *Lichtheimia hyalospora*, *Lichtheimia ramosa*, *Mucor ambiguus*, *Mucor circinelloides* und *Parasitella parasitica* besitzen zwischen fünf und sieben TSP1-ähnliche Sequenzen. *Backusella circina* hat mit neun TSP1-ähnlichen Proteinen die meisten Sequenzen dieser Art.

Mit Hilfe der TSP1-Sequenz aus *Mucor mucedo* wurden für alle betrachteten Zygomyceten TSP1-Sequenzen vom Typ1 identifiziert. Für *Rhizomucor pusillus* fand sich unter den BLAST-Ergebnissen eine zweite TSP1-ähnliche Sequenz (NCBI: BAM29042), die als Xylose-Reduktase annotiert ist. Eine BLAST-Suche mit dieser neuen TSP1-Variante führte zur Identifizierung weiterer TSP1-ähnlicher Sequenzen in den anderen Zygomyceten.

Abbildung 3.15 zeigt unter anderem die Strukturmodelle zweier TSP1-Varianten aus *Backusella circina*, *Mucor circinelloides* und *Rhizomucor pusillus*. Der Unterschied dieser beiden TSP1-Varianten zeigt sich an dem Auftreten/Fehlen einer kurzen Helix am Ende der Sequenz (Abbildung 3.13, Region 4).

Die beiden dargestellten TSP1-Varianten unterscheiden sich in der Zugänglichkeit der Bindetasche für das Trisporoid. Eine Variante ist weiter geöffnet, als die andere. Abbildung 3.16 zeigt, wie sich der Abstand zwischen den beiden flexiblen Bereichen Region 2 und Region 3 (Abbildung 3.13) auf die Zugänglichkeit der Bindetaschen für NADP und Trisporoid auswirkt.

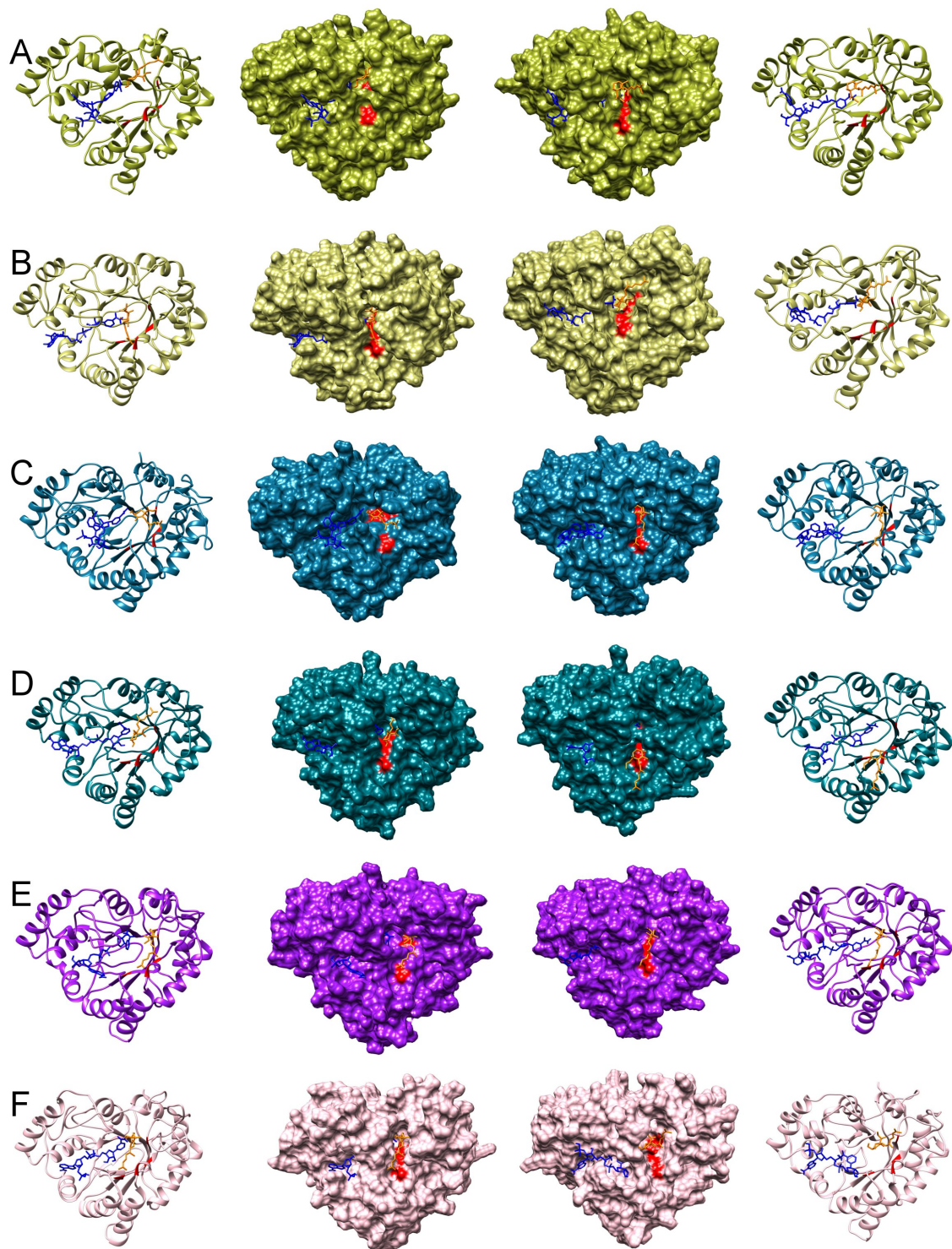


Abbildung 3.15: Modellierte TSP1-Strukturen für je zwei TSP1-Varianten aus (A) *Backusella circina* (251273 und 263122), (B) *Blakeslea trispora* (455083 und 482625), (C) *Lichtheimia hyalospora* (152197 und 207418), (D) *Mucor circinelloides* (CBS 277.49) (150877 und 155290), (E) *Rhizomucor pusillus* (BAN58735 und BAM29042) und (F) *Rhizopus microsporus* (248410 und 206793). Alle Proteinstrukturen beinhalten das beste Modell des Protein-Ligand-Dockings mit NADP (blau) und einem Trispurorid (orange). Das aktive Zentrum der Enzyme ist rot markiert.

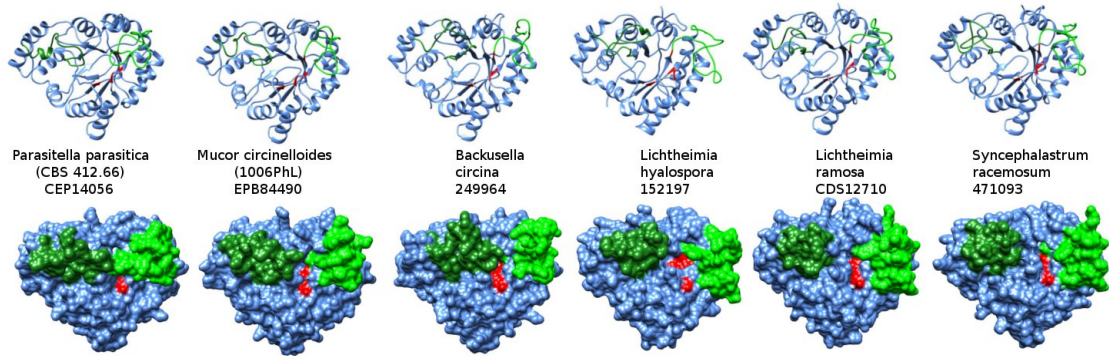


Abbildung 3.16: Modellierte TSP1-Strukturen veranschaulichen die Flexibilität der Regionen 2 (hellgrün) und 3 (dunkelgrün) und die damit verbundene Schließung der Bindetaschen. Das aktive Zentrum der Enzyme ist rot markiert.

TSP1-Proteine in homothallischen Zygomyceten

Homothallische Zygomyceten sind in der Lage, aus einem Myzel heraus sowohl (+)- als auch (–)-Kreuzungstyp-spezifische Zygophoren zu entwickeln. Es werden für die Entwicklung sexueller Strukturen keine zwei Individuen unterschiedlichen Kreuzungstyps benötigt, wie bei heterothallischen Pilzen (siehe Abschnitt 1.1.3). Die Anzahl der TSP1-Proteine schwankt in den betrachteten homothallischen Zygomyceten zwischen zwei und sechs. Abbildung 3.17 zeigt alle identifizierten TSP1-Proteine in *Mucor heterogamus*, *Mycotypha africana* und *Syncephalastrum racemosum*. Die modellierten Proteinstrukturen und Ligandenkonformationen unterscheiden sich nicht von denen in heterothallischen Zygomyceten.

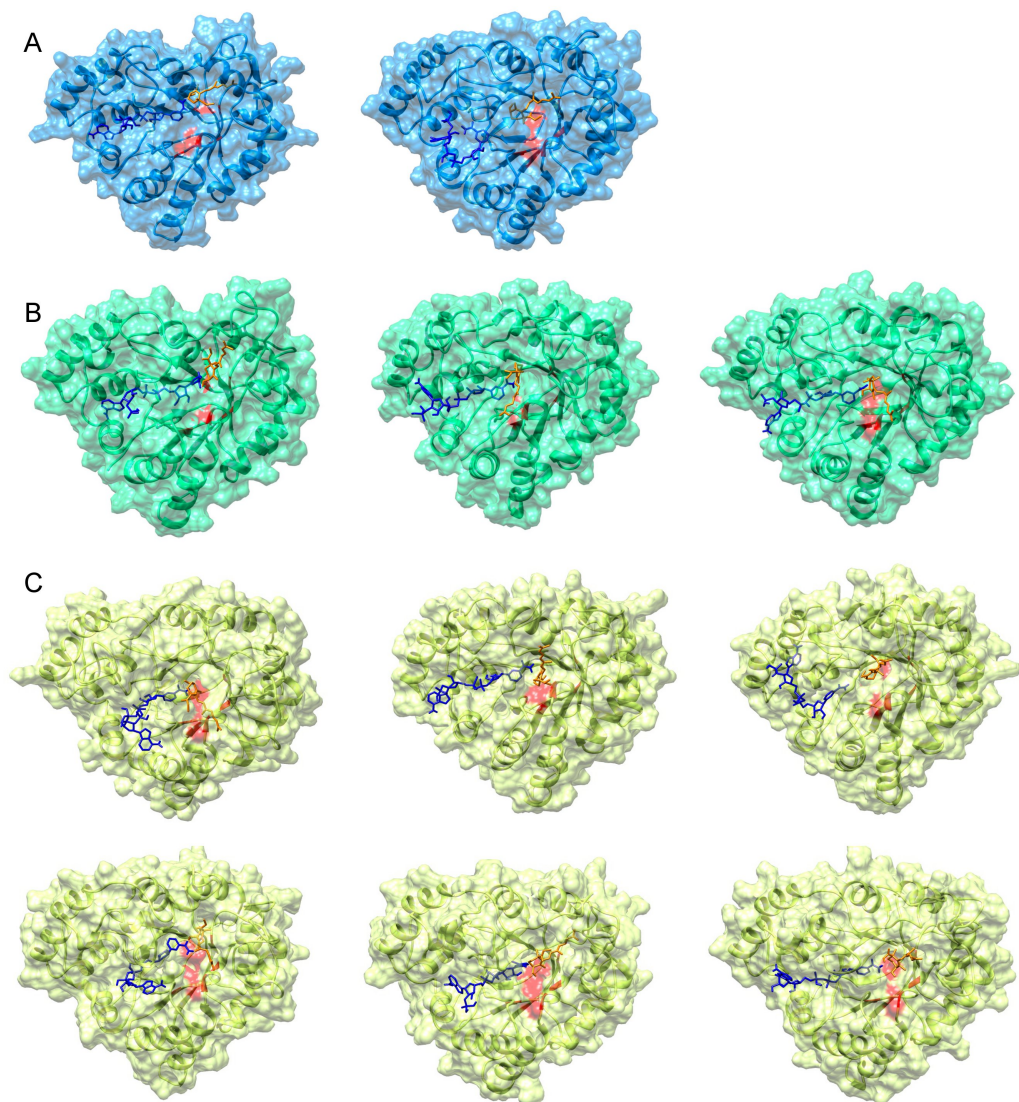


Abbildung 3.17: Modellierte TSP1-Strukturen für TSP1-Varianten aus den homothallischen Zygomyceten (A) *Mucor heterogamus* (524745 und 490332), (B) *Mycotypha africana* (816372, 875367 und 812904) und (C) *Syncephalastrum racemosum* (erste Reihe: 471093, 477980, 490367; zweite Reihe: 499703, 556430, 522060). Alle Proteinstrukturen beinhalten das beste Modell des Protein-Ligand-Dockings mit NADP (blau) und einem Trisporoid (orange). Das aktive Zentrum der Enzyme ist rot markiert.

TSP1-ähnliche Proteine außerhalb der Zygomyceten

Für den arbuskulären Mykorrhizapilz *Rhizophagus irregularis* DAOM 181602 wurde in der UniProt-Datenbank ein TSP1-ähnliches Protein identifiziert. Die modellierte Proteinstruktur wurde in *PLoS ONE* publiziert (Halary et al., 2013). Diese Proteinsequenz (UniProt: G8I0D9) ist als Xylose-Reduktase annotiert. In der JGI MycoCosm-Datenbank findet sich nur eine unvollständige Sequenz. Diese Sequenz (JGI: 350292) ist das erste Gen auf dem Scaffold 28269 und es fehlt das erste Exon am Anfang des Gens. Abbildung 3.18 zeigt ein Sequenzalignment mit der TSP1-Sequenz aus *Mucor mucedo*. Die Aminosäuren des aktiven Zentrums und der Bindetaschen für Substrat und Cosubstrat stimmen bis auf drei Positionen überein. Die Identität dieser beiden Sequenzen liegt bei 47 %.

```

Rhiid_G8I0D9  MSDNPTVKLNPTGQPMPLVGVMWVKVPNDKATDLVVEAFKLG YRLVDCASDYGNEKEIGI
Mucmu_Q01213  -MSTDYLTNRTGDKMPIRGFGCKIDTKDCEETVYQAIKTGYRLFDGACDYGNEVEVGR
               ..  :.* ** : ** : * . ** : .... : * :.* * ** : * .***** *:
               ..  :.* ** : ** : * . ** : .... : * :.* * ** : * .***** *:

Rhiid_G8I0D9  GLKKAFFDAGIVKREDIFVTSKLNNTNHARKHVRQAVERTLRDLQLDHLDLYLMHFTALK
Mucmu_Q01213  GINKAINEGLVKREDLFIVTKLNNTFHSKKHVRALFDRQLKDTGLEFYFDLYLIHFPVPLQ
               * :.* : * :.* ** : * :.* ** : * :.* ** : * :.* ** : * :.* ** :
               * :.* : * :.* ** : * :.* ** : * :.* ** : * :.* ** :

Rhiid_G8I0D9  YVDPEVRYPAEWYYPNKKVPIENIPIQETWQAMEELVDAGLVKNIGISNCCAGLIIDL
Mucmu_Q01213  YVDPATVYPGWYVGDAK-SLQFEQSPIHECWAELEKIVDAGLARNIGVANFNCQAILDL
               **** . ** * * * . : * : * * * * * : * :.* ** : * :.* ** :
               **** . ** * * * . : * : * * * * * : * :.* ** : * :.* ** :

Rhiid_G8I0D9  LRYARIKPSVLOIEHNPLYLTQETLINYVKSQGMVGTGYSNFGNLSYVDIIPKAKTAPILF
Mucmu_Q01213  LTYARIKPAVLQIELHPYLPQERLVKWKVKEQGQITAYSSHGFTSYVDLTESGKYTSLL
               * * * * * : * * * * * . * * * * * : * :.* ** : * :.* ** :
               * * * * * : * * * * * . * * * * * : * :.* ** : * :.* ** :

Rhiid_G8I0D9  DQPIIKELASSKNKSPAQIVLKWCVQRQIAIVEPKSNNHDLAQNKLQDFELTQEELHKI
Mucmu_Q01213  EHASVKSVDKHNVTGQVLLRWALDREFAVIPKSVNAGRMKANLEILDIKLD AEDNKTL
               : : . * : * : * * . * : * : * : * : * : * : * : * : * :
               : : . * : * : * * . * : * : * : * : * : * : * : * : * :

Rhiid_G8I0D9  SSLNINLRFNDPAD--FDLPFD
Mucmu_Q01213  DSLKTNQRFNDPMTYGFGLPLFD
               . * : * * * * * * * * *
               . * : * * * * * * * * *

```

Abbildung 3.18: Sequenzalignment des TSP1-Proteins aus *Mucor mucedo* und einem TSP1-ähnlichen Protein aus *Rhizophagus irregularis*.

Rot markierte Aminosäuren bilden das aktive Zentrum der Enzyme. Blau markierte Aminosäuren sind an der Bindung des Cofaktors NADP beteiligt. Orange markierte Aminosäuren sind an der Bindung des Trisoroid-Liganden beteiligt.

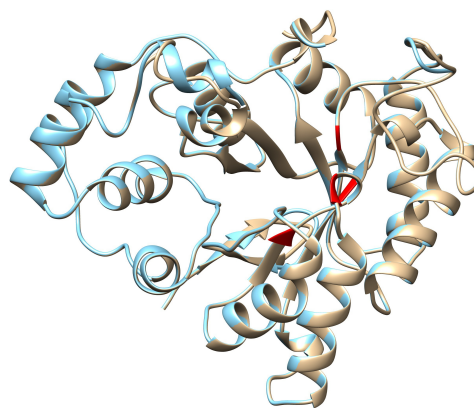


Abbildung 3.19: Überlagerung der TSP1-Proteinstruktur aus *Mucor mucedo* (blau: Q01213) und der Struktur eines TSP1-ähnlichen Proteins aus *Rhizophagus irregularis* (braun: G8I0D9).

3.2.2 4-Dihydrotrispurin-Dehydrogenase (TSP2)

Mit Hilfe einer BLAST-Suche mit der bereits identifizierten TSP2-Sequenz aus *Mucor mucedo* (UniProt: B9DR41) wurden weitere TSP2-Sequenzen für fast alle Zygomyceten in der JGI MycoCosm-Datenbank (Stand 07.2016) identifiziert. Mit Ausnahme von *Absidia repens* NRRL 1336, für die keine TSP2-Sequenzen gefunden werden konnten. Für *Backusella circina* FSU 941 (3), *Blakeslea trispora* NRRL 2456 (6), *Hesseltinella vesiculosa* NRRL 3301 (1), *Lichtheimia hyalospora* FSU 10163 (1), *Mucor circinelloides* CBS 277.49 (6), *Mucor heterogamus* NRRL 1489 (5), *Mycotypha africana* NRRL 2978 (2), *Parasitella parasitica* NRRL 2501 (2), *Phycomyces blakesleeanus* NRRL 1555 (4), *Rhizopus microsporus* ATCC 52813 (2), *Rhizopus oryzae* RA 99-880 (5), *Syncephalastrum racemosum* NRRL 2496 (1), *Thamnidium elegans* NRRL 2467 (1) und *Umbelopsis ramanniana* AG (2) wurden insgesamt 41 TSP2-Sequenzen identifiziert. Hinzukommen weitere 13 TSP2-Sequenzen aus Zygomyceten, die in NCBI identifiziert wurden. Diese stammen aus *Lichtheimia corymbifera* FSU 9682 (1), *Lichtheimia ramosa* FSU 6197 (1), *Mucor ambiguus* NBRC 6742 (4), *Mucor circinelloides* 1006PhL (5) und *Rhizomucor pusillus* NBRC 4578 (2). Auch in den von uns sequenzierten Zygomyceten *Absidia glauca* CBS 101.48 RVII-324 met⁻ (2) und *Parasitella parasitica* CBS 412.66 NGI-315 ade⁻ (2) wurden mehrere TSP2-Sequenzen identifiziert. In Klammern steht die Anzahl an TSP2-Sequenzen, die mindestens 95 % der Länge eines typischen TSP2-Proteins (240 Aminosäuren) aufweisen. In einigen Zygomyceten gibt es weitere TSP2-ähnliche Proteine, die aufgrund ihrer Länge oder ihres aktiven Zentrums nicht als TSP2 für die Analysen berücksichtigt wurden. Die Anzahl an TSP2-Sequenzen einer Spezies variiert deutlich. Einige Zygomyceten, wie zum Beispiel *Hesseltinella vesiculosa* und *Lichtheimia ramosa*, besitzen nur ein TSP2. Andere Zygomyceten, wie zum Beispiel *Blakeslea trispora*, besitzen überraschenderweise bis zu sechs TSP2- oder TSP2-ähnliche Proteine. Die Sequenzidentität der TSP2-Proteine liegt im Durchschnitt bei 53 %. Abbildung 3.22 zeigt einen Neighbor-Joining-Baum, der mit Clustal Omega (Sievers et al., 2011) erstellt wurde. Alle 58 TSP2-Sequenzen aus Zygomyceten sind darin enthalten. Es ist deutlich erkennbar, dass sich die verschiedenen TSP2-Sequenzen, anders als die betrachteten TSP1-Sequenzen in Abbildung 3.9, öfter nach dem Organismus gruppieren, aus dem sie stammen.

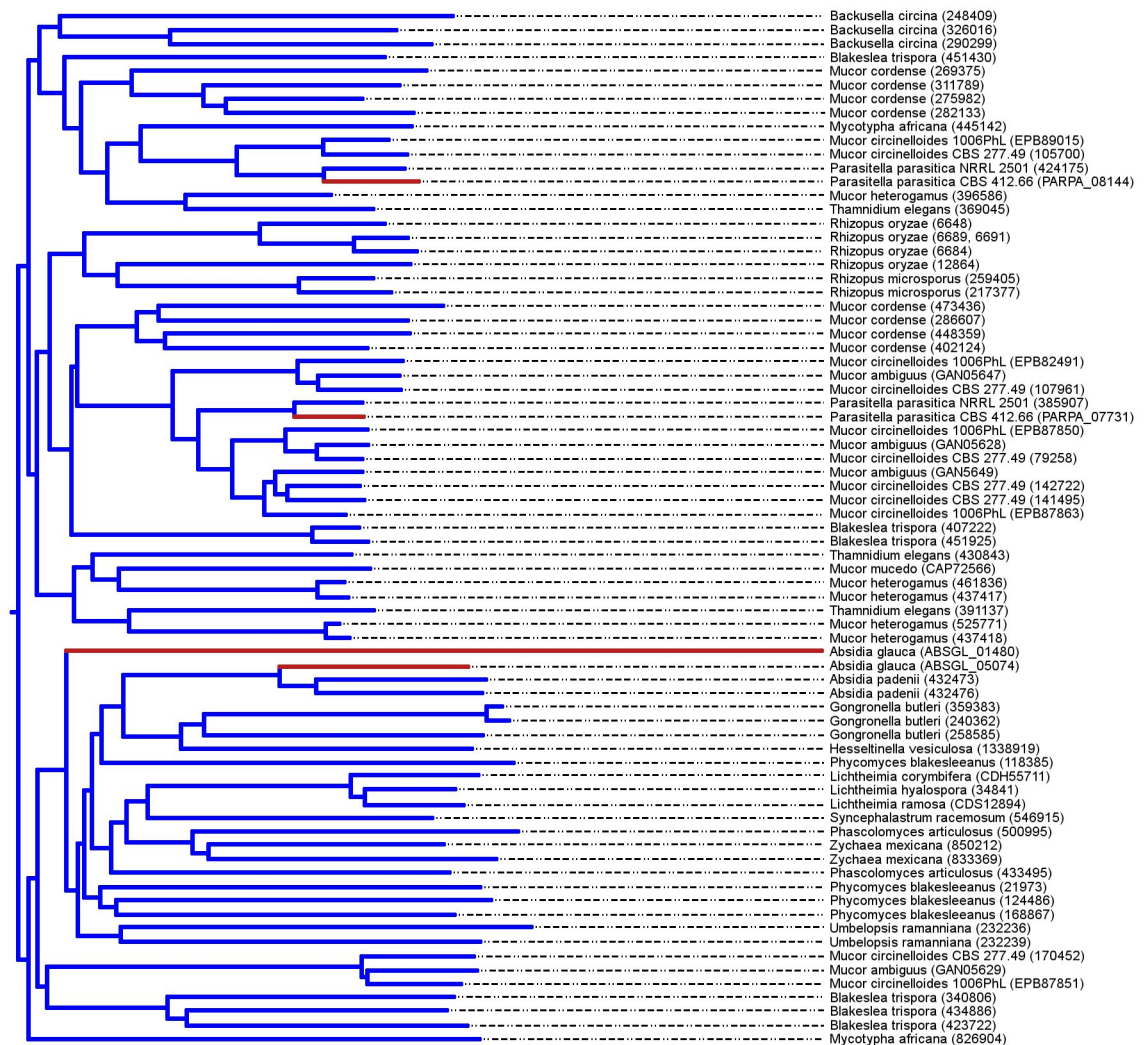


Abbildung 3.22: TSP2- und TSP2-ähnliche Proteine aus 26 verschiedenen Zygomyceten. Die Sequenzen aus *Absidia glauca* und *Parasitella parasitica* sind rot markiert.

Nr.	Protein-ID	Datenbank	Länge (AS)	MG (kDa)	TGxxxGxG (Cosubstratbindung)	YxxxK (Aktives Zentrum)
<i>Absidia glauca</i> (CBS 101.48 RVII-324 met-, (-)-Kreuzungstyp)						
Seq1	SAL99460	NCBI	237	25,25	11-18	159-163
Seq2	SAL96112	NCBI	244	26,18	8-15	152-156
<i>Absidia repens</i> (NRRL 1336, (+)-Kreuzungstyp)						
Kein TSP2 identifiziert						
<i>Backusella circina</i> (FSU 941, (+)-Kreuzungstyp)						
Seq1	326016	JGI / MycoCosm	238	25,52	7-14	160-164
Seq2	248409	JGI / MycoCosm	241	26,13	8-15	163-167
Seq3	290299	JGI / MycoCosm	239	25,86	7-14	161-165
<i>Blakeslea trispora</i> (NRRL 2456, (+)-Kreuzungstyp)						
Seq1	407222	JGI / MycoCosm	240	25,55	8-15	162-166
Seq2	340806	JGI / MycoCosm	242	26,21	9-16	164-168
Seq3	423722	JGI / MycoCosm	241	26,29	9-16	163-167
Seq4	434886	JGI / MycoCosm	241	26,25	11-18	163-167
Seq5	451430	JGI / MycoCosm	237	25,44	8-15	159-163
Seq6	451925	JGI / MycoCosm	240	25,81	8-15	162-166
<i>Hesseltinella vesiculosa</i> (NRRL3301, (-)-Kreuzungstyp)						
Seq1	1338919	JGI / MycoCosm	234	25,16	8-15	156-160
<i>Lichtheimia corymbifera</i> (FSU 9682, (-)-Kreuzungstyp)						
Seq1	CDH55711	NCBI	232	25	8-15	154-158
<i>Lichtheimia hyalospora</i> (FSU 10163, (-)-Kreuzungstyp)						
Seq1	34841	JGI / MycoCosm	232	24,8	8-15	154-158
<i>Lichtheimia ramosa</i> (FSU 6197, (-)-Kreuzungstyp)						
Seq1	CDS12894	NCBI	232	24,73	8-15	154-158
<i>Mucor ambiguus</i> (NBRC 6742, (+)-Kreuzungstyp)						
Seq1	GAN05628	NCBI	240	25,75	8-15	162-166
Seq2	GAN05629	NCBI	243	26,01	9-16	165-169
Seq3	GAN05647	NCBI	241	26,16	8-15	163-167
Seq4	GAN05649	NCBI	241	25,72	8-15	163-167
<i>Mucor circinelloides</i> (CBS 277.49, (-)-Kreuzungstyp)						
Seq1	79258	JGI / MycoCosm	240	25,64	8-15	162-166
Seq2	105700	JGI / MycoCosm	240	25,61	8-15	162-166
Seq3	107961	JGI / MycoCosm	241	26,11	8-15	163-167
Seq4	141495	JGI / MycoCosm	232	24,69	8-15	154-158
Seq5	142722	JGI / MycoCosm	240	25,68	8-15	162-166
Seq6	170452	JGI / MycoCosm	243	26,2	9-16	165-169

Tabelle 3.8: Liste aller TSP2-Sequenzen aus Zygomyceten. Bei mehreren TSP2-Sequenzen wurde nur die erste in Sequenz- und Strukturalignments berücksichtigt. Unterstrichene Sequenzen wurden experimentell als TSP2 identifiziert.

Nr.	Protein-ID	Datenbank	Länge (AS)	MG (kDa)	TGxxxGxG (Cosubstratbindung)	YxxxK (Aktives Zentrum)
<i>Mucor circinelloides</i> (1006PhL, (+)-Kreuzungstyp)						
Seq1	EPB87850	NCBI	240	25,55	8-15	162-166
Seq2	EPB82491	NCBI	241	26,2	8-15	163-167
Seq3	EPB87851	NCBI	243	26,32	9-16	165-169
Seq4	EPB87863	NCBI	238	25,51	8-15	160-164
Seq5	EPB89015	NCBI	240	25,69	8-15	162-166
<i>Mucor heterogamus</i> (NRRL 1489, homothallisch)						
Seq1	461836	JGI / MycoCosm	240	26,06	8-15	162-166
Seq2	437417	JGI / MycoCosm	240	26,08	8-15	162-166
Seq3	525771	JGI / MycoCosm	240	26,11	8-15	162-166
Seq4	437418	JGI / MycoCosm	230	24,99	8-15	162-166
Seq5	396586	JGI / MycoCosm	240	25,68	8-15	162-166
<i>Mucor mucedo</i> (CBS109.16, (-)-Kreuzungstyp)						
Seq1	<u>CAP72566</u>	NCBI	240	25,81	8-15	162-166
<i>Mycotypha africana</i> (NRRL 2978, homothallisch)						
Seq1	445142	JGI / MycoCosm	240	25,5	8-15	162-166
Seq2	826904	JGI / MycoCosm	242	26,44	11-18	164-168
<i>Parasitella parasitica</i> (CBS 412.66 NGI-315 ade-, (+)-Kreuzungstyp)						
Seq1	CEP13616	NCBI	240	25,72	8-15	162-166
Seq2	CEP13991	NCBI	240	25,82	8-15	162-166
<i>Parasitella parasitica</i> (NRRL 2501, (-)-Kreuzungstyp)						
Seq1	385907	JGI / MycoCosm	240	25,89	8-15	162-166
Seq2	424175	JGI / MycoCosm	240	25,57	8-15	162-166
<i>Phycomyces blakesleeanus</i> (NRRL 1555, (+)-Kreuzungstyp)						
Seq1	124486	JGI / MycoCosm	236	25,07	8-15	158-162
Seq2	21973	JGI / MycoCosm	231	25,06	8-15	153-157
Seq3	118385	JGI / MycoCosm	231	24,98	11-18	153-157
Seq4	168867	JGI / MycoCosm	232	24,76	8-15	154-158
<i>Rhizopus microsporus</i> (ATCC 52813, (+)-Kreuzungstyp)						
Seq1	259405	JGI / MycoCosm	234	25,37	7-14	156-160
Seq2	217377	JGI / MycoCosm	234	25,47	7-14	156-160
<i>Rhizopus oryzae</i> (RA 99-880, (+)-Kreuzungstyp)						
Seq1	6648	JGI / MycoCosm	234	25,27	7-14	156-160
Seq2	6684	JGI / MycoCosm	240	26,07	7-14	156-160
Seq3	6689, 6691	JGI / MycoCosm	240	26,23	7-14	156-160
Seq4	12864	JGI / MycoCosm	245	26,88	7-14	156-160

Tabelle 3.8: (Fortsetzung)

Nr.	Protein-ID	Datenbank	Länge (AS)	MG (kDa)	TGxxxGxG (Cosubstratbindung)	YxxxK (Aktives Zentrum)
<i>Syncephalastrum racemosum</i> (NRRL 2496, homothallisch)						
Seq1	546915	JGI / MycoCosm	232	24,74	7-14	154-158
<i>Thamnidium elegans</i> (NRRL 2467, (+)-Kreuzungstyp)						
Seq1	430843	JGI / MycoCosm	240	25,49	8-15	162-166
Seq2	369045	JGI / MycoCosm	240	25,58	8-15	162-166
Seq3	391137	JGI / MycoCosm	237	25,84	8-15	159-163
<i>Umbelopsis ramanniana</i> (AG, (-)-Kreuzungstyp)						
Seq1	232239	JGI / MycoCosm	235	25,22	9-16	157-161
Seq2	232236	JGI / MycoCosm	235	25,75	9-16	157-161

Tabelle 3.8: (Fortsetzung)

Abbildung 3.23 zeigt ein multiples Sequenzalignment der in Zygomyceten identifizierten TSP2-Proteine. Die Aminosäuren des aktiven Zentrums, Tyrosin und Lysin, sind in allen TSP2- und TSP2-ähnlichen Proteinen konserviert. Die Aminosäuren, die an der Bindung des Cosubstrats NADP beteiligt sind (Ellenberger et al., 2013), stimmen im Alignment an den meisten Positionen überein. Über die an der Bindung des Trisporoid-Liganden beteiligten Aminosäuren können keine exakten Angaben, so wie bei TSP1-Proteinen, gemacht werden. Die Gruppe der kurzkettigen Dehydrogenasen ist sehr heterogen in ihrer Sequenz und Funktion (Kallberg et al., 2002). Die bei der Modellierung als Vorlage verwendeten Proteine (Tabelle 2.1) binden eine Vielzahl verschiedener Liganden, wie zum Beispiel Alkohole, Zucker, Steroide und aromatische Verbindungen. Sie liefern daher kein einheitliches Bild der Substrat-Bindetasche.

Die typische Tertiärstruktur von kurzkettigen Dehydrogenasen (Kavanagh et al., 2008) besteht aus einer charakteristischen Rossmann-Faltung (Rossmann et al., 1974). Diese setzt sich aus einem zentralen siebensträngigen β -Faltblatt und sechs α -Helices zusammen. Drei dieser Helices liegen auf der einen Seite des β -Faltblatts, die anderen auf der gegenüberliegenden Seite.

Da die Kernstruktur bei TSP2-Proteinen und kurzkettigen Dehydrogenasen gut konserviert ist, ähnelt sich die Struktur der modellierten TSP2-Proteine, obwohl unterschiedliche Vorlagen (Tabelle 2.1) für die einzelnen Proteine ausgewählt wurden (Abbildung 3.24 (A)). In (Abbildung 3.24) (B) sind die Bereiche, die eine größere Variabilität aufweisen, im Detail dargestellt. Region 3 beinhaltet bei dem TSP2-Protein aus *Mucor mucedo* zwei zusätzliche Helices (*a1* und *a2*) in der Schleife, die für die Spezifität der Substratbindung verantwortlich ist. Diese Schleife ist bei *Mucor mucedo* 16 Aminosäuren länger, als bei den übrigen betrachteten TSP2-Proteinen aus Zygomyceten.

Ein multiples Struktur-Alignment einiger TSP2-Proteine (Abbildung 3.25) zeigt die Unterschiede in der Struktur der TSP2-Proteine.

Mycotypha africana_445142	---MAKTYIVTTCASRGICLEFVKQLLRGDIVFAAARNPDSSKGLQALSNGNEKLHLVQLD
Rhizopus microsporus_259405	----MSLYVITTCASRGICLEFVKQLSAKNDNTVFACARNPDQAEELKKLVNDKKVFSIKLD
Rhizopus oryzae_6648	----MVSYYITTCASRGICLEFVKQLSARGDTVFACARNPDKAEGLQKLVNDKKVFSIKLD
Backusella circina_326016	----MSVYVITTCASRGICLEFVKQLAEAGKTVFACARNPDQKSEKLSVLIDNKKHVFQIKLD
Thamnidium elegans_430843	---MTSTYVITTCASRGICLEFVKQLSSKNGTVFALARNPDASEGLQKLVNDKKVFAFKLD
Mucor heterogamus_461836	---MPLTYVVTTCASRGICLEFVKQLISQNDHKVFALARNPDSSSEGLKALVDNKNVFAVKVD
Mucor mucedo_CAP72566	---MSLTYYITTCASRGICLEFVNQIAAKGHTVIALARNPDASEGLQKLVNDKKVFSVQLE
Blakeslea trispora_407222	---MSLTYYITTCASRGICLEFVQRQIAAKGNTVFACARNPDASEGLQKLVNDKKVFSVKLD
Parasitella parasitica (NRRL 2501)_385907	---MSSTYVVTTCASRGICLEFVRQLAKGNTVFACARNPDGNADKLKELVDDKHVFAIKFD
Parasitella parasitica (CBS 412.66)_CEP13616	---MPSTYVVTTCASRGICLEFIKQLAEKNGTVFACARNPDGNADKLKELVDDKHVFAIKLD
Mucor circinelloides (I006PhL)_EPB87850	---MSLTYYVVTTCASRGICLEFVNQLAAKNGTVIACARNPDGSAAGLQKLVNDKNHFAVKMD
Mucor ambiguus_GAN05628	---MSLTYYVVTTCASRGICLEFVSQLAAKNGTVIACARNPDNAAGLQKLVNDKNHFAVKMD
Mucor circinelloides (CBS 277.49)_79258	---MSLTYYVVTTCASRGICLEFVSQLAAKNGTVIACARNPDNAAGLQKLVNDKNHFAVKMD
Phycomyces blakesleeanus_124486	---MTLTYYLVTCASRGICLEFVKQLASAGNTVIATARNIEASSGLKALAQD-NVILTTL
Absidia glauca_SAL99460	MSAQQLTYLVTCASRGICLEFVKQLSAKGHIATARDPTKAEALVSLIDYEKVGIVGVALD
Hesseltinella vesiculosa_1338919	---MAKTVLITTCSSRGICLEFVRQLINKGEIVIATARNLSSATELMALVDNKSIVFAVALD
Umbelopsis ramanniana_232239	---MAKTTFYITTCASRGICLEFVSQLKRAGHVVIATARKPENSEELQKLVNDKQVFSIALD
Syncephalastrum racemosum_546915	----MSTYVITTCASRGICLEFVKQLDKGHTVIACARPEKSDALQALVDNQRIHSLKLD
Lichtheimia corymbifera_CDH55711	---MGLTWVITTCSTRGICAEFVKQLKNGKDTVACARNPDSSDALKALVDGKQVHAVKMD
Lichtheimia hyalospora_34841	---MGLTWVITTCSTRGICAEFVKQLKNGKDTVACARNPDGSEALQALVDGKQVHAVKMD
Lichtheimia ramosa_CDS12894	---MGLTWVITTCSTRGICAEFVKQLKNGKDTVACARNPDGSDALKSLVDGKQVHAVMD
	:*:***:* *: *: *: *:***: * *: *
	TGxxxGxG → klassische kurz-kettige Dehydrogenase
Mycotypha africana_445142	ALSEESIKAAAAEIEKKAPGVDVLINNAGICGEMGLDFEQTSKNELVVKFFETNVAAVNE
Rhizopus microsporus_259405	TTNEQSLKGAVDEISKHAPNGIDVLINNAGIALSRK-TIEETGKDELMKTFETNVAGVVE
Rhizopus oryzae_6648	TACEKSIKAEVDEISKLAPEGDFDVLINNAGIAGAYT-APEQTSKKEILEIFETNVLAINE
Backusella circina_326016	TTTDSIQTAARKEIDEKAPNGVNVVLINNAGISGDAQYDVLTPESDYVNVVFTNNVGVSN
Thamnidium elegans_430843	ALPESSIKNAVEEVTKASPEGIDYVLINNAGISGERGYGISNSPGEEYTKVFNTNVAVSN
Mucor heterogamus_461836	TTSEKSIKNAVEEVSKNAPEGVDILINNAGISGSGSMNYTPEDTPKNEYIDVFQTNVGVSD
Mucor mucedo_CAP72566	TTTEKSIKSAVEEASKISPOGIDVLINNAGISGSRDGNIRKTDKEDYIHVFNTNVAVAD
Blakeslea trispora_407222	TLKSDSLKAAVEYIEKHTSEGDVLINNAGIGGSSASTPDKVTENLLKAFETNVCGTSD
Parasitella parasitica (NRRL 2501)_385907	TTTLESIKSGVEEISKHAPQGDVLINNAGIGGNSKNPQTLLLEEDLINFQTNVGVGSN
Parasitella parasitica (CBS 412.66)_CEP13616	TTTPEKSIKAAVEEVSKHAPQGDVLINNAGIGGNSKSPETLLLEEDLITFQTNVGVSN
Mucor circinelloides (I006PhL)_EPB87850	TTTLESIKSAAEEVAKYAPQGDVLINNAGIGGEKATKPEDLFESDLNNIFQTNVGVGSN
Mucor ambiguus_GAN05628	TTTNYDSIKSAAVEVSKYAPQGDVLINNAGIGGEKSTRPENLLEEDLNTFQTNVGVSN
Mucor circinelloides (CBS 277.49)_79258	TTTYESIKTAAQEVSKYAPQGDVLINNAGIGGEKSTRPENLLEEDLNTFQTNVGVSN
Phycomyces blakesleeanus_124486	TTGSESIIKAEVEFVSKVAPNGIDVLINNAGATGDMSLDVLNCTPEDYRKAFFINVGTS
Absidia glauca_SAL99460	TTTQSSVKAFAFEKIKTIAPHGIDVLINNAGISGPASEALTATADQLVDFTKNVGVTSN
Hesseltinella vesiculosa_1338919	TTTCDMSVKAAVEEIGRIAPTGDVLINNAGIAGATDATFLNTDKDDFMKVFETNVGVTSN
Umbelopsis ramanniana_232239	TTTVKSVKAAVEEIKVNEIAPNGIDVLINNAGVSTDRSHSLTTPPEAMLDVFKTNVGVTVI
Syncephalastrum racemosum_546915	TTVCADSIKAAVEEKIHQLAPGGIDVLINNAGIAGERGQDTLTTTAEDYINVFKTNVGTSD
Lichtheimia corymbifera_CDH55711	TTTVASVKAATDKISLAPEGIDVLINNAGVGT-PQTEPLTIQPEAYTNDFTNVGVTSI
Lichtheimia hyalospora_34841	ITTTVASVKAATDKISLAPEGIDILLNNSGVGT-MQTDLTLPPEAYINDFTNVGVTSN
Lichtheimia ramosa_CDS12894	VTTVASVKAATDKISQLAPEGIDILINNSGVGT-MKTDVLDLPPEAYINDFTNVGVTSN
	*::: . : *: .*:***: . * ** ..
Mycotypha africana_445142	VSKAFVPIRLKRKGAGCVKKILNMSSSLGSISSMEDATGYGFGVAYCVSRAGTNMLTRAMS
Rhizopus microsporus_259405	VTNAFIPLLRKRKGDKDTHKKIVNISSLLGSVGLCDSDS---KLAAYRVRSKAVNMVTKLQS
Rhizopus oryzae_6648	VTNAFIPLLRKRKGRDVRVKKILNISSILGSVELVNF---PETASVRISSKALNMLTKLQA
Backusella circina_326016	VVQGFPLLRKSG-DGVKKIMNMSSILGSIAMDEGSGWGFSGSYCVSKAAVNMMTKMKS
Thamnidium elegans_430843	VTHAFIPLLRKRKGQEKTKKILNISSILGSVALIGEVNPGSGSTYSVSKALNMLTKLFA
Mucor heterogamus_461836	VTQAFIPLLRKRKGQDKTKKILNMSSILASIAIDIVINSNRGSAYSVSKAALNMLTKMFA
Mucor mucedo_CAP72566	VTTAFIPLLRKRGIESTKKIVNMSSGLGSIARINTINPTGRGSAYSVSKAALNMLTKMFA
Blakeslea trispora_407222	STQAFIPLLRKRKGDNVKKIVNISSILGSIKINEINGSGGVPEYNISKAALNMLTKHTTA
Parasitella parasitica (NRRL 2501)_385907	VTKAFIPPLQKRGRGDHVKKILNISSILGSIELINQIAPTGSQPPAYNISKAALNMVVKMLS
Parasitella parasitica (CBS 412.66)_CEP13616	VTKAFIPPLQKRGRGQEHVKKILNISSILGSIELINQVSPGSGQPPAYNISKAALNMVVKMLS
Mucor circinelloides (I006PhL)_EPB87850	VTKVFLPPLQKRGRGDHVKKIVNISSILGSIELINQIAPTGGQPSYNISKAALNMVTKMLA
Mucor ambiguus_GAN05628	VTKEFLPPLQKRGRGDHVKKIVNISSILGSIELINQISPTGGQPPAYNISKAALNMVVKMLA
Mucor circinelloides (CBS 277.49)_79258	VTKEFLPPLQKRGRGDHVKKIVNISSILGSIELINQVSPGSGQPPAYNISKAALNMVVKMLA
Phycomyces blakesleeanus_124486	VTQAMFPLLRKLGK---TRKIVNLCSVLGSIERTENGNAWGFMSYAVSKAEAGMLTKGFS
Absidia glauca_SAL99460	VTQALPPLLRQKD---TRKIINISSIVGSIKLNH---GATMPYGVRSKAGENYLTKALS
Hesseltinella vesiculosa_1338919	VSOQATDLRLKRD---TRIIFNISSVLGSIELNPG---QGHKTAYSVSKAENMLTRFSA
Umbelopsis ramanniana_232239	VTQRFPVPLLRKKT---DRKIPTISSILGSVGLTNY---AGTIAFNVSKAAVNMYIKELS
Syncephalastrum racemosum_546915	VTRALPVLRLQKN---TRHVFNISSILGSVENTT---AGGNVSRVSKAAVNMLSRATA
Lichtheimia corymbifera_CDH55711	VIQALPVLRLKRD---TRRIINISSIMSGMECST---ESFNPSYRVSKAAVNMLTRVFA
Lichtheimia hyalospora_34841	VIQALPVLRLKRD---TRRIINISSILGSMEITT---DGSNPSYRVSKAAVNMLTRAVA
Lichtheimia ramosa_CDS12894	VIQALPVLRLKRD---TRRIINISSILGSIENTT---EGHGSYRVSKAAVNMLTRAVA
	:*: : : : .*: :*: * :*: : *
	klassische kurz-kettige Dehydrogenase ← YxxxK
Mycotypha africana_445142	SKLGKENFIVYSSHPGMYCTDMGGGAPVTEASISGMLARLDSITAEDNGKFFDFESKTIW
Rhizopus microsporus_259405	LQFAENFIVYSAHPGMYCTDMGGGAHAITPDVSIKGMLSKLESATSKDNGGFFSYTGESMEW
Rhizopus oryzae_6648	MQLAKENIIVYASHPGMYCTDMGGGENDIYADESISGQLAKLDSATAADAGKLINYKGEVLNY
Backusella circina_326016	NALAKENIIVYASHPGMYCTDMGSSKAPTRTEDSISKQLVLRLDLAKDNGAYIDFEGNHLNW
Thamnidium elegans_430843	NSLAKENIIVYAAHPGMYCTDMGGDSAPVEPKDSIAGMVKVLQSVTAEOQSGLDFEGKQLPW
Mucor heterogamus_461836	NQLASENFIVYASHPGMYCTDMGGENAPVERHDSIAGMLKIIDLTPEKNGAFIDFQKGLKW
Mucor mucedo_CAP72566	NQLASENFIVYASHPGMYCTDLGGEDAPYDKKDSIAGMLKIIENLTPEQNGSFIFEGDELFW
Blakeslea trispora_407222	HVLAKENFIVYASHPGMYCTDMGGESAPVEPKDSVAGQLAKLESATSKDNGGFFDFEGNTLPW
Parasitella parasitica (NRRL 2501)_385907	NNLAKENFIVYSSHPGMYCTDMGGESAPVEVEDSISKQLAKLESATSKDNGGFFDFEGKALPW
Parasitella parasitica (CBS 412.66)_CEP13616	NNLAKENFIVYSSHPGMYCTDLGGESAPVEAEDSISKQLAKLESATSKDNGGFFDFEGKALPW
Mucor circinelloides (I006PhL)_EPB87850	NNLAKDNFIVYSSHPGMYCTDLGGESAPVQAVDSIRGQLAKLENATSKDNGGFFDFEGKTLPW
Mucor ambiguus_GAN05628	NNLAKDNFIVYSSHPGMYCTDLGGESAPVQAVDSIRGQLAKLESATSKDNGGFFDFEGKTLPW
Mucor circinelloides (CBS 277.49)_79258	NNLAKDNFIVYSSHPGMYCTDLGGESAPVQAVDSIRGQLAKLESATSKDNGGFFDFEGKTLPW
Phycomyces blakesleeanus_124486	CALKDEDDFTIVAVHPGMYCTDMGGKNGELTTQTSVEGMLNCVYKMTTKDNGTFFDYAGTILPW
Absidia glauca_SAL99460	ASLAEKFIVVAVHPGMYCTDMGGQKAPLKAESIAGVEQIEKLTQKNGMYLFTGESLFW
Hesseltinella vesiculosa_1338919	NGLKNESFIVSVHPGMYCTDMGGSQHLSPESESITGMIEQLYKLQRSQNGTFFDYKGDALPW
Umbelopsis ramanniana_232239	EDLAKDNFTILAIHPGMYCTDMGGASAPLKAPESSISGMLKVFQDQATEDTGSFKDYSGKTLPW
Syncephalastrum racemosum_546915	QGLASENFIVVAVHPGMYCTDMGGGQAPVTPQSGIRGMLSVFDKVTQADNGTFFDYKGETLPW
Lichtheimia corymbifera_CDH55711	AQLKDEGFIIVCVHPGMYCTDMGTTHAPLTTDQSIGGMLSAFDKLTASNGTFFDYKGETLPW
Lichtheimia hyalospora_34841	QQLKDEGFIIVVAVHPGMYCTDMGTTHAPLTTDQSVGMLAGFDKLNKESNGTFFDYKGETLPW
Lichtheimia ramosa_CDS12894	GQLKDEGFIIVVAVHPGMYCTDMGTDRAPLTTDQSVGGMLSAFDKLNKESNGTFFDYKGETLPW
	: : : : .**** *:***: . *:***: . * : : : :

Abbildung 3.23: Sequenzalignment der TSP2-Proteine in Zygomyceten.

Rot markierte Aminosäuren bilden das aktive Zentrum der Enzyme. Blau markierte Aminosäuren sind an der Bindung des Cofaktors NADP beteiligt.

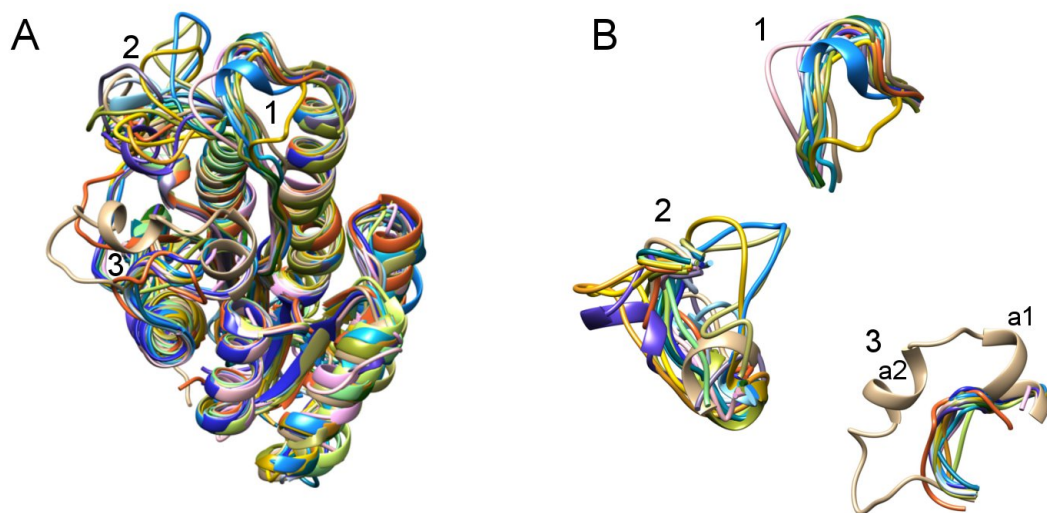
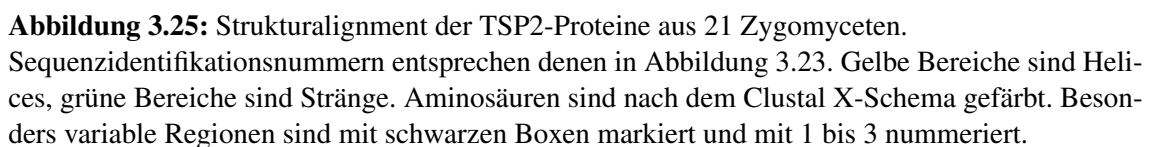


Abbildung 3.24: (A) Überlagerung der Proteinstrukturen von TSP2-Sequenzen aus 21 Zygomyceten. (B) Die Ziffern 1-3 bezeichnen variable Regionen der Strukturen. Diese Bereiche werden hier im Detail dargestellt. Die Farben entsprechen denen in Abbildung 3.25.



Zygomyceten mit einem TSP2-Protein

Hesseltinella vesiculosa, *Lichtheimia corymbifera*, *Lichtheimia hyalospora*, *Lichtheimia hyalospora* und *Mucor mucedo* besitzen nur ein TSP2-Protein. Für *Mucor mucedo* wurde das komplette Genom noch nicht sequenziert. Die Verwandtschaft zu anderen Zygomyceten wie *Mucor circinelloides* lässt vermuten, dass es weitere TSP2-Proteine in diesem Pilz gibt. Abbildung 3.26 zeigt die modellierten Proteinstrukturen und die Ergebnisse aus dem Protein-Ligand-Docking für diese TSP2-Varianten. In TSP2-Proteinen gibt es für das Trisoroid nur eine sterisch mögliche Position, um mit dem Cosubstrat in Kontakt zu kommen und eine Reaktion zu ermöglichen. Durch einen engen Tunnel auf der gegenüberliegenden Seite des Proteins gelangt der Ligand zum aktiven Zentrum und zum Cosubstrat NADP.

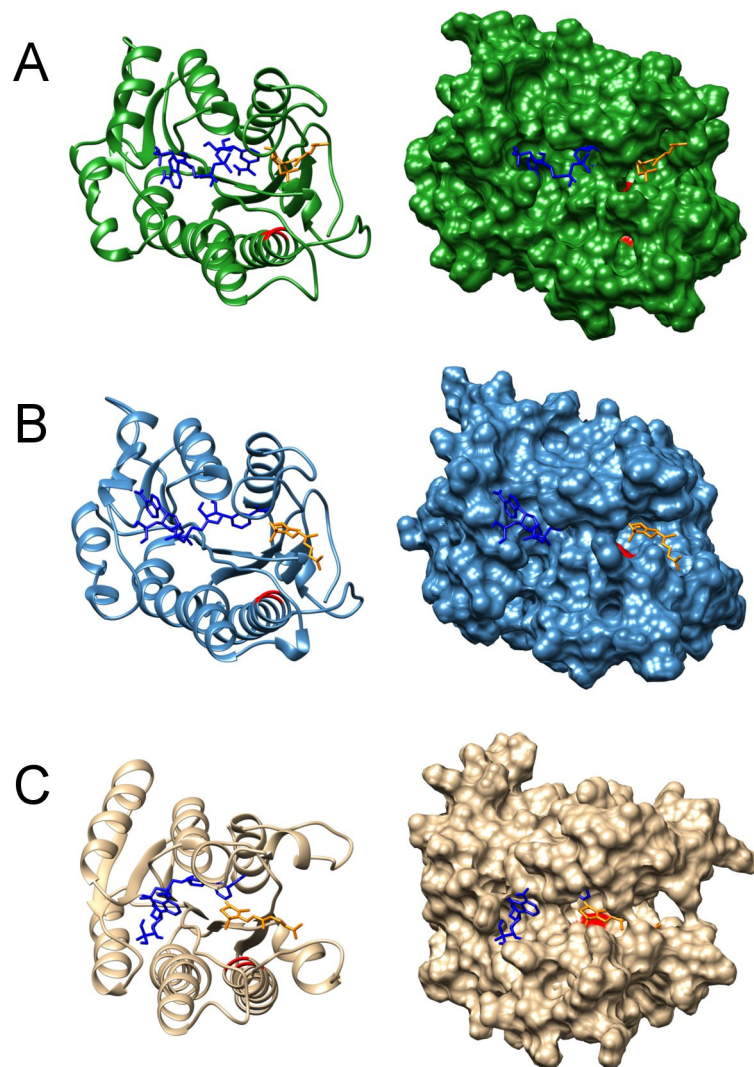


Abbildung 3.26: Modellierte TSP2-Strukturen für (A) *Hesseltinella vesiculosa*, (B) *Lichtheimia hyalospora* und (C) *Mucor mucedo*. Alle Proteinstrukturen beinhalten das beste Modell des Protein-Ligand-Dockings mit NADP (blau) und einem Trisoroid (orange). Das aktive Zentrum der Enzyme ist rot markiert.

Zygomyceten mit bis zu sechs TSP2-Proteinen

Absidia glauca, *Backusella circina*, *Mucor ambiguus*, *Mucor circinelloides* 1006PhL, *Rhizopus microsporus* und *Rhizopus oryzae* besitzen zwischen zwei und fünf TSP2-Sequenzen. *Blakeslea trispora* und *Mucor circinelloides* CBS 277.49 haben mit sechs TSP2-ähnlichen Proteinen die meisten Sequenzen dieser Art. Abbildung 3.27 zeigt jeweils zwei ausgewählte TSP2-Varianten und die Ergebnisse aus dem Protein-Ligand-Docking für die modellierten Proteinstrukturen.

Im Gegensatz zu TSP1 ist das NADP-Molekül in einigen TSP2-Proteinen nicht linear angeordnet. Es passt sich einer verzweigten Bindetasche an (Abbildung 3.27 Variante 1 in (C), (E) und (G)). Das stellt kein Problem dar, da alle chemischen Bindungen, die an der Reaktion beteiligt sind, rotierbar sind. Diese ungewöhnlich flexible Struktur des Co-substrats hat keinen negativen Einfluss auf den Abstand zwischen NADP und Trisporoid. Eine Reaktion kann in dieser Positionierung stattfinden.

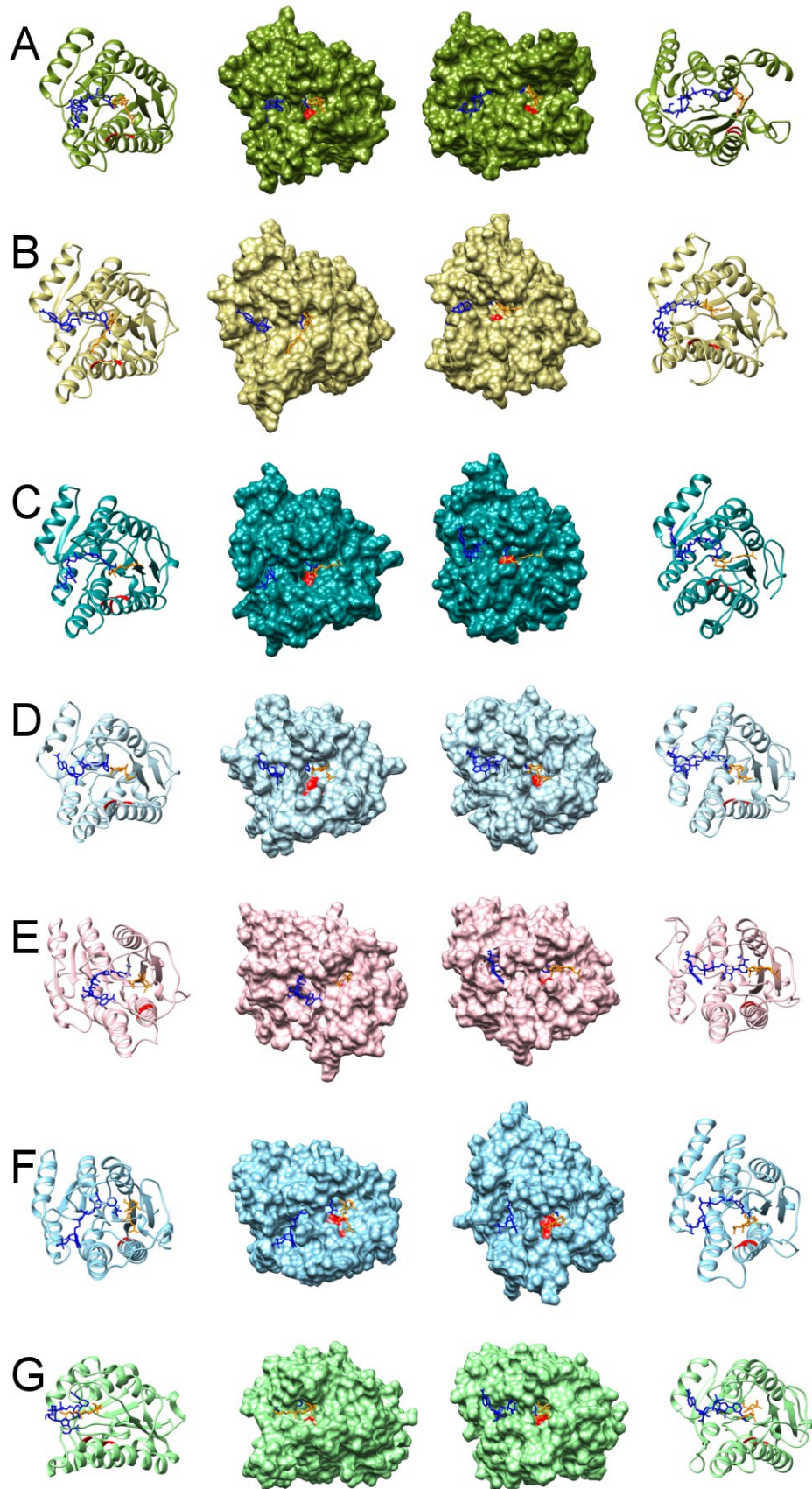


Abbildung 3.27: Ausgewählte Modelle für TSP2-Proteine aus (A) *Backusella circina* (326016, 248409), (B) *Blakeslea trispora* (434486, 451925), (C) *Mucor circinelloides* (CBS 277.49) (79258, 170452), (D) *Phycomyces blakesleeanus* (124486, 21973), (E) *Rhizopus microsporus* (217377, 259405), (F) *Thamnidium elegans* (430843, 369045) und (G) *Umbelopsis ramanniana* (232236, 232239). Alle Proteinstrukturen beinhalten das beste Modell des Protein-Ligand-Dockings mit NADP (blau) und einem Trispurin (orange). Das aktive Zentrum der Enzyme ist rot markiert.

TSP2-Proteine in homothallischen Zygomyceten

Abbildung 3.28 zeigt die modellierten Proteinstrukturen und Ergebnisse aus dem Protein-Ligand-Docking für alle identifizierten TSP2-Proteine aus *Mucor heterogamus*, *Mycotypha africana* und *Syncephalastrum racemosum*. Die modellierten Proteinstrukturen und Ligandenkonformationen unterscheiden sich nicht von denen in heterothallischen Zygomyceten. Auch in homothallischen Zygomyceten gibt es stark gefaltete NADP-Moleküle (Abbildung 3.28 (B)).

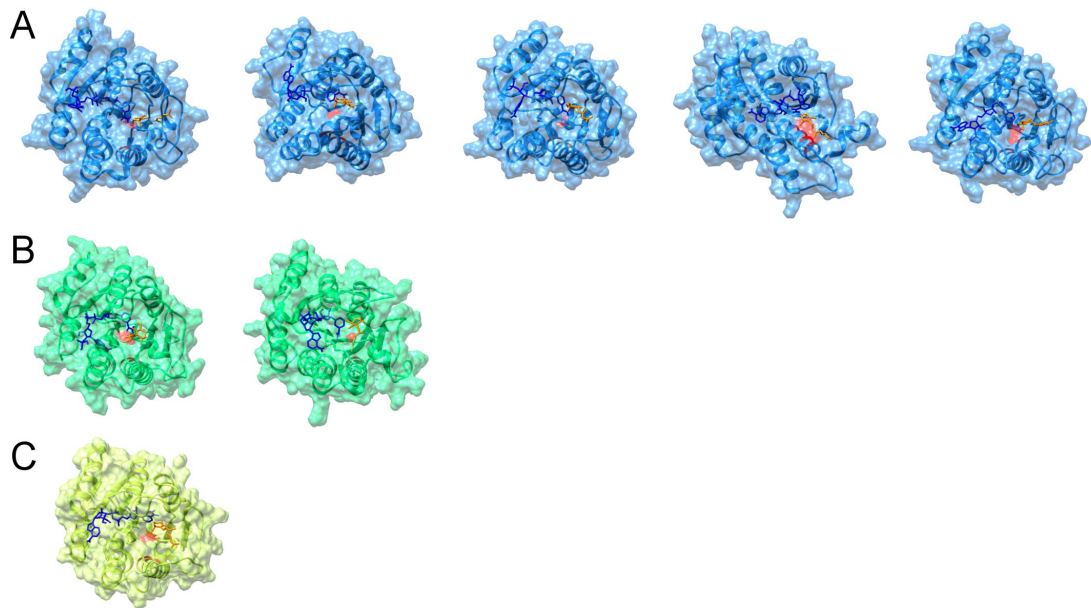


Abbildung 3.28: Modellierte TSP2-Strukturen für TSP2-Varianten aus den homothallischen Zygomyceten (A) *Mucor heterogamus* (461836, 437417, 525771, 437418, 396586), (B) *Mycotypha africana* (445142, 826904) und (C) *Syncephalastrum racemosum* (546915). Alle Proteinstrukturen beinhalten das beste Modell des Protein-Ligand-Dockings mit NADP (blau) und einem Trispurinoid (orange). Das aktive Zentrum der Enzyme ist rot markiert.

3.2.3 Carotin-Oxygenase (TSP3)

Der erste Schritt in der Produktion von Trisporsäure ist die oxidative Spaltung von β -Carotin. Diese Reaktion wird von einer Carotin-Oxygenase (TSP3) durchgeführt. Mit Hilfe einer BLAST-Suche mit der bereits identifizierten TSP3-Sequenz aus *Blakeslea trispora* (UniProt: A5HXP5) wurden weitere TSP3-Sequenzen für alle Zygomyceten in der JGI MycoCosm-Datenbank und NCBI (Stand 07.2016) identifiziert. Im Gegensatz zu TSP1 und TSP2 gibt es für TSP3 in jedem Zygomyceten nur ein Protein.

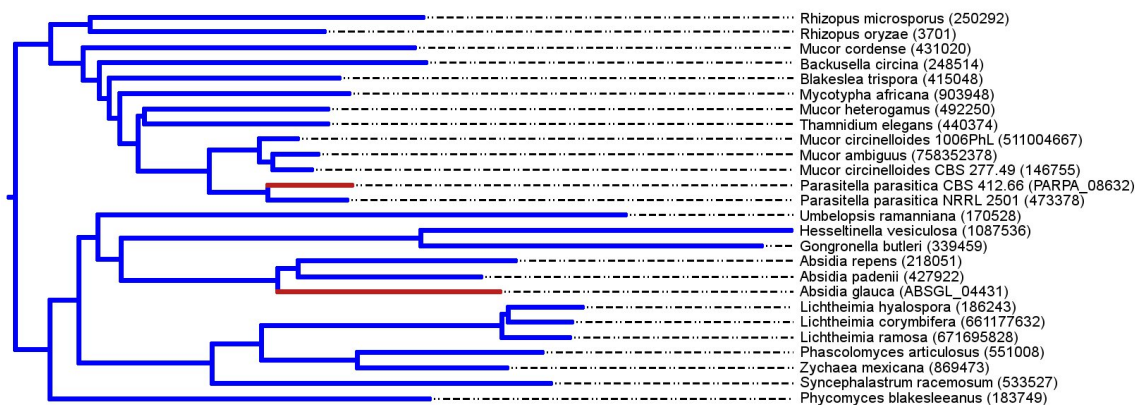


Abbildung 3.29: TSP3-Proteine aus 26 verschiedenen Zygomyceten. Die Sequenzen aus *Absidia glauca* und *Parasitella parasitica* sind rot markiert.

Tabelle 3.9 enthält alle identifizierten TSP3-Sequenzen. Sie haben eine Länge von 607 – 646 Aminosäuren und ein Molekulargewicht (MG) von circa 70 kDa.

Abbildung 3.30 zeigt ein multiples Sequenzalignment der in Zygomyceten identifizierten TSP3-Proteine. Das Eisenion im Zentrum von TSP3 wird von vier Histidinen positioniert. Diese sind in allen Carotinoid-Oxygenasen konserviert (Sui et al., 2013). Die Sequenzidentität der TSP3-Proteine liegt im Durchschnitt bei 73 %.

Nr.	Protein-ID	Datenbank	Länge (AS)	Molekular- gewicht (kDa)	Eisenionbindung
<i>Absidia glauca</i> (CBS 101.48 RVII-324 met-, (-)-Kreuzungstyp)					
Seq1	SAL98866	NCBI	616	69,42	H-216, H-283, H-351, H-602
<i>Absidia repens</i> (NRRL 1336, (+)-Kreuzungstyp)					
Seq1	218051	JGI / MycoCosm	646	73,14	H-216, H-294, H-362, H-632
<i>Backusella circina</i> (FSU 941, (+)-Kreuzungstyp)					
Seq1	248514	JGI / MycoCosm	627	70,57	H-220, H-292, H-360, H-613
<i>Blakeslea trispora</i> (NRRL 2456, (+)-Kreuzungstyp)					
Seq1	415048	JGI / MycoCosm	627	70,91	H-220, H-291, H-359, H-613
<i>Hesseltinella vesiculosa</i> (NRRL3301, (-)-Kreuzungstyp)					
Seq1	1087536	JGI / MycoCosm	623	70,34	H-219, H-282, H-350, H-609
<i>Lichtheimia corymbifera</i> (FSU 9682, (-)-Kreuzungstyp)					
Seq1	CDH59423	NCBI	624	70,37	H-218, H-286, H-354, H-610
<i>Lichtheimia hyalospora</i> (FSU 10163, (-)-Kreuzungstyp)					
Seq1	186243	JGI / MycoCosm	624	70,56	H-218, H-286, H-354, H-610
<i>Lichtheimia ramosa</i> (FSU 6197, (-)-Kreuzungstyp)					
Seq1	CDS10547	NCBI	624	70,47	H-218, H-286, H-354, H-610
<i>Mucor ambiguus</i> (NBRC 6742, (+)-Kreuzungstyp)					
Seq1	GAN05477	NCBI	624	70,19	H-218, H-290, H-358, H-610
<i>Mucor circinelloides</i> (CBS 277.49, (-)-Kreuzungstyp)					
Seq1	146755	JGI / MycoCosm	624	70,25	H-218, H-290, H-358, H-610
<i>Mucor circinelloides</i> (1006PhL, (+)-Kreuzungstyp)					
Seq1	EPB86031	JGI / MycoCosm	624	70,04	H-218, H-290, H-358, H-610
<i>Mucor heterogamus</i> (NRRL 1489, homothallisch)					
Seq1	492250	JGI / MycoCosm	628	71,06	H-220, H-292, H-360, H-614
<i>Mycotypha africana</i> (NRRL 2978, homothallisch)					
Seq1	903948	JGI / MycoCosm	628	71,44	H-220, H-292, H-360, H-614
<i>Parasitella parasitica</i> (CBS 412.66 NGI-315 ade-, (+)-Kreuzungstyp)					
Seq1	CEP14452	NCBI	620	70,04	H-214, H-286, H-354, H-606
<i>Parasitella parasitica</i> (NRRL 2501, (-)-Kreuzungstyp)					
Seq1	473378	JGI / MycoCosm	624	70,3	H-218, H-290, H-358, H-610
<i>Phycomyces blakesleeanus</i> (NRRL1555, (+)-Kreuzungstyp)					
Seq1	183749	JGI / MycoCosm	628	70,17	H-220, H-292, H-360, H-614
<i>Rhizopus microsporus</i> (ATCC52813, (+)-Kreuzungstyp)					
Seq1	250292	JGI / MycoCosm	625	70,76	H-220, H-292, H-360, H-611
<i>Rhizopus oryzae</i> (RA 99-880, (+)-Kreuzungstyp)					
Seq1	3701	JGI / MycoCosm	626	70,73	H-220, H-292, H-360, H-612
<i>Syncephalastrum racemosum</i> (NRRL 2496, homothallisch)					
Seq1	533527	JGI / MycoCosm	616	69,51	H-220, H-286, H-354, H-602
<i>Thamnidium elegans</i> (NRRL 2467, (+)-Kreuzungstyp)					
Seq1	440374	JGI / MycoCosm	628	70,98	H-220, H-292, H-360, H-614
<i>Umbelopsis ramanniana</i> (AG, (-)-Kreuzungstyp)					
Seq1	170528	JGI / MycoCosm	607	68,18	H-215, H-284, H-352, H-593

Tabelle 3.9: Liste aller TSP3-Sequenzen aus Zygomyceten.

Hesseltinella_vesiculosa_1087536
Umbelopsis_ramanniana_170528
Syncephalastrum_racemosum_533527
Lichtheimia_hyalospora_186243
Lichtheimia_corymbifera_661177632
Lichtheimia_ramosa_671695828
Absidia_glauca_SAL98866
Absidia_repens_218051
Phycomyces_blakesleeana_183749
Rhizopus_microsporus_250292
Backusella_circina_248514
Rhizopus_oryzae_3701
Blakeslea_trispora_415048
Mycotypha_africana_903948
Mucor_heterogamus_492250
Thamnidium_elegans_440374
Mucor_circinelloides_(1006PhL)_511004667
Mucor_ambiguus_758352378
Mucor_circinelloides_(CBS_277.49)_146755
Parasitella_parasitica_(CBS_412.66)_CEP14452
Parasitella_parasitica_(NRRL_2501)_473378

[illegible]

Abbildung 3.30: Sequenzalignment der TSP3-Proteine aus Zygomyceten. Orange markierte Aminosäuren sind an der Bindung des zentral gelegenen Eisenions beteiligt.

Hesselтинella_vesiculosa_1087536
Umbelopsis_ramanniana_170528
Syncephalastrum_racemosum_533527
Lichtheimia_hyalospora_186243
Lichtheimia_corymbifera_661177632
Lichtheimia_ramosa_671695828
Absidia_glauca_SAL98866
Absidia_repens_218051
Phycomyces_blakesleeana_183749
Rhizopus_microsporus_250292
Bacularella_circina_248514
Rhizopus_oryzae_3701
Blakeslea_trispora_415048
Mycotypha_africana_903948
Mucor_heterogamus_492250
Thamnidium_elegans_440374
Mucor_circinelloides_(1006PhL)_511004667
Mucor_ambiguus_758352378
Mucor_circinelloides_(CBS_277.49)_146755
Parasitella_parasitica_(CBS_412.66)_CEP14452
Parasitella_parasitica_(NRRL_2501)_473378

```

-----SDGQVDVLDKRVAPARFVRLDKHTDNKLLKHEMKPSEVRRFRFLANVPDYFIKEEDRD
-----VDFFDAEKKVPKAPFVI--KKPAQKGKANNIAPSQVRRYRLGNVPVSNVSRNV--
-----SNE--MDLHAVQPARFVI--KKPAQKGKPGSIAPQVRRYRLGNVPQSDSL
-----APGSKGVNESKVKPFRFVV--KKKAQKGKPGNKIAPAQVRRYRLGNVQVPSDENQ
-----APGKGVNESKVKPFRFVV--KKKAQKGKPGNKIAPAQVRRYRLGNVQVPSDENQ
-----APGKGVNESKVKPFRFVV--KKKAQKGKPGNKIAPAQVRRYRLGNVQVPSDENQ
-----SSSPPTTKIDLDKVVPARYVHIKNTTEGKTDHAIAPSQVRRYRLANVPDQF
-----SNSNPTIKTSKVDLDKVPARYVHIKNTTEGKSDHAIAPSQVRRYRLGNVPDQGLQQQFPQ
-----PTG--FDASKVEPARFKI--KRHTDDKDKNISPSQLRYYRLGNVPVSNNAPE--
-----PSGRLNDKVKPARFV--KATANKTEHKIAPSQVRRYRLGNVPVSK--PDT--
-----PSGKLDLDNVKPARFV--NKKTNDKKDHQIKPSRFCFRRLGNVPVAK--PDE--
-----PSGKLDLDLVQPACFI--KNTDGKHEKHSIPSQVRRYRLGNVPVSK--PGA--
-----PTGKLDLDKVKPARFR--DKLTDHKTEHQIAPSQVRRYRLGNVPVAK--PDA--
-----PTGKLDLDQVQPARFV--DKNTHHKKEHQIAPSQVRRYRLGNVPVITE--SSA--
-----PTGKLDLDLVQPARFV--DKSYQKKEHQIAPSQVRRYRLGNVPVSK--PGA--
-----PTGKLDLDQVQPARFV--DKNTDKKADHQIAPSQVRRYRLGNVPVSK--PDA--
-----PTGKLDLDLVQPARFV--DKNTAHKEEHQIAPSQVRRYRLGNVPVAK--PGA--
-----PTGKLDLDLVQPARFV--DKNTAHKEEHQIAPSQVRRYRLGNVPVAK--PGA--
-----PTGKLDLDLVQPARFV--DKNTAHKKEHQIAPSQVRRYRLGNVPVAK--PGA--
-----PTGMDLDQVQPARFV--DKNTAHKKEHQIAPSQVRRYRLGNVPVAK--FSD--
-----PTGQMDLDQVKPARYIF--DKNTAHKKEHQIAPSQVRRYRLGNVPVAK--FGD--

```

Abbildung 3.30: (Fortsetzung)

Abbildung 3.30: (Fortsetzung)

TSP3-Proteine in heterothallischen Zygomyceten

Aus der Gruppe der Enzyme, die Carotin spalten, sind nur drei dreidimensionale Proteinstrukturen bekannt. Die Proteinstrukturen einer Apocarotinoid spaltenden Oxygenase (PDB: 2BIW), einer Retinoid-Isomerohydrolase (PDB: 3FSN) und einer 9-cis-Epoxy-carotinoid-Dioxygenase (PDB: 3NPE) wurden für die Modellierung der TSP3-Proteine verwendet. Abbildung 3.31 zeigt die Proteinmodelle einiger heterothallischer Zygomyceten. Die Größe der Öffnung, durch die das Carotin in das Enzym gelangt, unterscheidet sich zwischen den einzelnen TSP3-Proteinen. Bei TSP3 aus *Phycomyces blakesleeanus* ist das aktive Zentrum nicht von außen sichtbar.

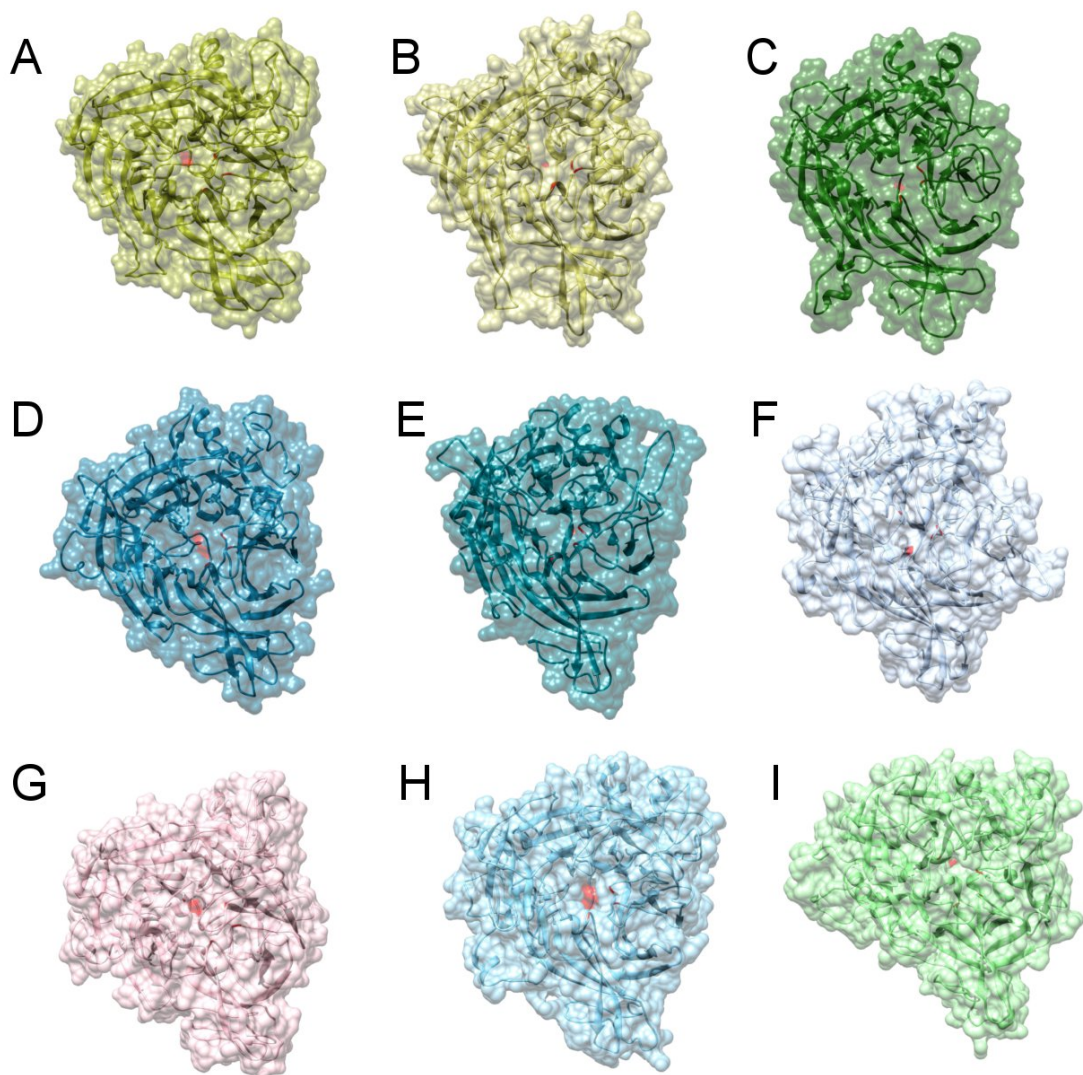


Abbildung 3.31: Modellierte TSP3-Proteine aus (A) *Backusella circina*, (B) *Blakeslea trispora*, (C) *Hesseltnella vesiculosa*, (D) *Lichtheimia hyalospora*, (E) *Mucor circinelloides*, (F) *Phycomyces blakesleeanus*, (G) *Rhizopus microsporus*, (H) *Thamnidium elegans* und (I) *Umbelopsis ramanniana*. Die Aminosäuren, die an der Bindung des zentral gelegenen Eisenions beteiligt sind, sind rot markiert.

TSP3-Proteine in homothallischen Zygomyceten

Abbildung 3.32 zeigt die Proteinmodelle der betrachteten homothallischen Zygomyceten. Die Größe der Öffnung, durch die das Carotin in das Enzym gelangt, unterscheidet sich zwischen den einzelnen TSP3-Proteinen. Bei TSP3 aus *Mycotypha africana* ist das aktive Zentrum nicht von außen sichtbar.

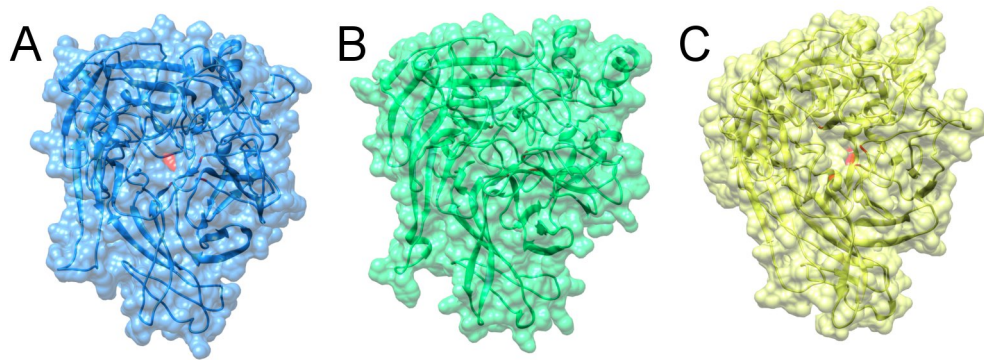


Abbildung 3.32: Modellierte TSP3-Proteine aus den homothallischen Zygomyceten (A) *Mucor heterogamus*, (B) *Mycotypha africana* und (C) *Syncephalastrum racemosum*. Die Aminosäuren, die an der Bindung des zentral gelegenen Eisenions beteiligt sind, sind rot markiert.

TSP3-Proteine in *Absidia*- und *Parasitella*-Stämmen

Die Sequenzidentität der TSP3-Proteine aus *Absidia glauca* und *Parasitella parasitica* liegt bei 71 %. Abbildung 3.33 zeigt die Proteinmodelle der TSP3-Proteine aus *Absidia glauca* und *Parasitella parasitica*. Die Öffnung, durch die das Carotin in das Enzym gelangt, ist bei *Absidia glauca* größer als bei *Parasitella parasitica*.

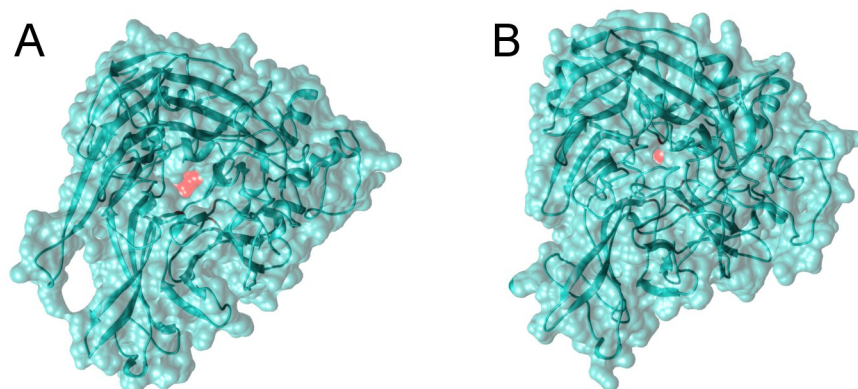


Abbildung 3.33: Modellierte TSP3-Proteine aus (A) *Absidia glauca* und (B) *Parasitella parasitica* CBS 412.66. Die Aminosäuren, die an der Bindung des zentral gelegenen Eisenions beteiligt sind, sind rot markiert.

3.3 Analyse der Ligandenbindung in TSP1- und TSP2-Proteinen

Die Spezifität der Bindung eines Trisporoid-Liganden wird in TSP1 und TSP2 über flexible Schleifenstrukturen vorgegeben. Posttranslationale Regulation, die mit der Bildung von Dimeren realisiert wird, verändert die Konformation der Substrat-Bindetaschen dieser Proteine und führt bei einigen Proteinkomplexen zur Inaktivierung des Enzyms. Diese Mechanismen erschweren die Untersuchung der Ligandenbindung in TSP1- und TSP2-Monomeren.

3.3.1 Ligandenbindung bei TSP1

Die Zugänglichkeit des aktiven Zentrums variiert deutlich zwischen den einzelnen TSP1-Isomeren. In AutoDock Vina werden zehn Modelle für ein Protein-Ligand-Docking ausgegeben. M1 ist dabei das Modell mit dem niedrigsten Energieniveau, das an erster Stelle der Rangfolge steht. Tabelle 3.10 zeigt, wie zugänglich das aktive Zentrum der Enzyme für Substrat und Cosubstrat ist. In dieser Tabelle sind die Ergebnisse des Protein-Ligand-Dockings für die betrachteten TSP1-Proteine aufgeführt. Modelle, die weiter hinten in der Reihenfolge der Ausgabe stehen, sind nicht automatisch schlechte Vorhersagen. Die Abweichung der freien Bindungsenergie der Moleküle kann zwischen den Modellen sehr gering sein, sodass keine echte Wertung der Platzierung vorliegen muss. Ein Beispiel hierfür sind die Modelle für eine freie Energie von -6,1 kcal/mol. Ein Modell aus *Absidia glauca* (NCBI: SAM06439), das diese Energie aufweist, liegt auf Platz 10 und ist somit das letzte Modell der Ausgabe. Ein Modell aus *Absidia repens* (JGI: 352134), das diese Energie aufweist, liegt dagegen auf Platz 4. Die niedrigste ausgegebene Energie für ein in einem TSP1-Protein modelliertes Trisporoid liegt bei -9,3 kcal/mol und stammt aus *Mucor heterogamus* (JGI: 524745). Die höchste ausgegebene Energie für ein in einem TSP1-Protein modelliertes Trisporoid liegt bei -4,5 kcal/mol und stammt aus *Backusella circina* (JGI: 282851).

Bei einigen TSP1-Proteinen sind die Bindetaschen für Substrat und Cosubstrat miteinander verbunden (NCBI: CDH51836 aus *Lichtheimia corymbifera*, JGI: 152197 aus *Lichtheimia hyalospora*, NCBI: GAN08850 aus *Mucor ambiguus*, JGI: 124728 aus *Phycomyces blakesleeanus*, JGI: 471093 aus *Syncephalastrum racemosum* und JGI: 241297 aus *Umbelopsis ramanniana*). Diese gemeinsame Binderegion liegt offen zugänglich auf der Proteinoberfläche und ist dadurch weniger spezifisch.

Einige TSP1-Proteine haben keine offen zugängliche Bindetasche für Trisporoide (JGI: 221028 aus *Backusella circina*, NCBI: CDS08007 aus *Lichtheimia ramosa*, NCBI: EPB81770 aus *Mucor circinelloides* 1006PhL und NCBI: CEP14056 aus *Parasitella parasitica* CBS 412.66). Bei ihnen handelt es sich entweder um inaktive TSP1-Isomere, die ausschließlich der Regulation anderer aktiver TSP1-Isomere dienen oder um Teile eines

Proteinkomplexes, der erst nach Anlagerung weiterer Partner aktiviert wird.

In vielen TSP1-Proteinen sind die Zugänge für Substrat und Cosubstrat zum aktiven Zentrum hin verengt. Die Ergebnisse des Protein-Ligand-Dockings zeigen, dass diese jedoch in den meisten Fällen trotzdem noch groß genug sind, um NADP und Trisporoid so nah aneinander herankommen zu lassen, dass eine Reaktion möglich ist.

Nr.	Protein-ID	Datenbank	Enzymaktivität (Zugänglichkeit des aktiven Zentrums)		AutoDock Vina Freie Bindungsenergie (kcal/mol)	
			NADP	Trisporoid	NADP	Trisporoid
Absidia glauca (CBS 101.48 RVII-324 met-, (-)-Kreuzungstyp)						
Seq1	SAL98187	NCBI	verengt	verengt	M6: -7,5	M2: -7,7
Seq2	SAL98947	NCBI	verengt	offen	M2: -8,5	M8: -6,4
Seq3	<u>SAM02260</u>	NCBI	offen	offen	M1: -9,8	M8: -6,7
Seq4	SAM06439	NCBI	verengt	verengt	M1: -7,8	M10: -6,1
Absidia repens (NRRL 1336, (+)-Kreuzungstyp)						
Seq1	375419	JGI / MycoCosm	verengt	verengt	M7: -4,1	M2: -5,5
Seq2	452045	JGI / MycoCosm	verengt	verengt	M5: -7,7	M2: -7,1
Seq3	384469	JGI / MycoCosm	offen	offen	M1: -8,2	M4: -7,1
Seq4	352134	JGI / MycoCosm	verengt	offen	M5: -8,3	M4: -6,1
Backusella circina (FSU 941, (+)-Kreuzungstyp)						
Seq1	251273	JGI / MycoCosm	verengt	verengt	M4: -7,6	M5: -8,3
Seq2	263122	JGI / MycoCosm	verengt	verengt	M3: -7,9	M1: -5,9
Seq3	249964	JGI / MycoCosm	offen	offen	M3: -9,0	M8: -6,1
Seq4	221028	JGI / MycoCosm	offen	geschlossen	M4: -9,2	M1: -5,6
Seq5	320949	JGI / MycoCosm	verengt	offen	M1: -9,2	M4: -7,0
Seq6	240807	JGI / MycoCosm	offen	offen	M3: -9,2	M1: -6,0
Seq7	321804	JGI / MycoCosm	offen	verengt	M5: -8,4	M5: -6,9
Seq8	254831	JGI / MycoCosm	offen	verengt	M3: -9,4	M6: -6,1
Seq9	282851	JGI / MycoCosm	verengt	offen	M10: -6,4	M3: -4,5

Tabelle 3.10: Ergebnisse des Protein-Ligand-Dockings mit TSP1-Proteinen aus Zygomyceten. Aufgelistet sind die besten Modelle für die Bindung des Cosubstrats NADP und eines Trisporoid-Liganden. Die Ziffer hinter dem M bezeichnet die Position des Modells in der Rangfolge der Ausgabe. Unterstrichene Sequenzen wurden experimentell als TSP1 identifiziert.

~: Die Bindetaschen für NADP und Trisporoid liegen offen zugänglich an der Oberfläche des Proteins und sind nicht voneinander getrennt, wie bei der Mehrheit der TSP1-Proteine.

Nr.	Protein-ID	Datenbank	Enzymaktivität (Zugänglichkeit des aktiven Zentrums)		AutoDock Vina Freie Bindungsenergie (kcal/mol)	
			NADP	Trisporoid	NADP	Trisporoid
Blakeslea trispora (NRRL 2456, (+)-Kreuzungstyp)						
Seq1	455083	JGI / MycoCosm	verengt	verengt	M2: -8,3	M6: -6,5
Seq2	482625	JGI / MycoCosm	verengt	offen	M5: -8,4	M6: -8,1
Hesseltinella vesiculosa (NRRL 3301, (-)-Kreuzungstyp)						
Seq1	304048	JGI / MycoCosm	verengt	offen	M1: -8,7	M9: -5,7
Lichtheimia corymbifera (FSU 9682, (-)-Kreuzungstyp)						
Seq1	CDH51040	NCBI	verengt	verengt	M1: -8,1	M8: -6,5
Seq2	CDH48981	NCBI	verengt	verengt	M2: -8,1	M4: -5,6
Seq3	CDH51836	NCBI	verengt ~	verengt ~	M4: -8,4	M6: -7,1
Seq4	CDH58265	NCBI	verengt	verengt	M1: -10,5	M5: -6,8
Seq5	CDH59815	NCBI	verengt	verengt	M1: -7,1	M2: -6,1
Seq6	CDH51044	NCBI	offen	offen	M1: -10,0	M9: -5,5
Seq7	CDH61395	NCBI	offen	verengt	M1: -8,1	M10: -5,6
Lichtheimia hyalospora (FSU 10163, (-)-Kreuzungstyp)						
Seq1	152197	JGI / MycoCosm	offen ~	offen ~	M4: -7,3	M6: -7,0
Seq2	207418	JGI / MycoCosm	verengt	verengt	M3: -7,2	M3: -6,7
Seq3	168802	JGI / MycoCosm	offen	offen	M7: -8,4	M4: -6,4
Seq4	197017	JGI / MycoCosm	verengt	verengt	M1: -10,7	M6: -7,2
Seq5	124793	JGI / MycoCosm	verengt	verengt	M1: -9,8	M6: -7,4
Seq6	165119	JGI / MycoCosm	verengt	offen	M1: -8,4	M9: -6,2
Seq7	213704	JGI / MycoCosm	offen	offen	M1: -8,3	M4: -6,3
Lichtheimia ramosa (FSU 6197, (-)-Kreuzungstyp)						
Seq1	CDS12710	NCBI	verengt	verengt	M3: -8,6	M9: -5,4
Seq2	CDS07919	NCBI	offen	verengt	M2: -9,7	M7: -6,3
Seq3	CDS03129	NCBI	offen	offen	M7: -8,0	M3: -7,1
Seq4	CDS09639	NCBI	offen	verengt	M1: -10,7	M8: -5,3
Seq5	CDS09007	NCBI	verengt	verengt	M2: -7,7	M1: -5,8
Seq6	CDS12708	NCBI	verengt	verengt	M1: -10,8	M1: -7,8
Seq7	CDS08007	NCBI	offen	geschlossen	M1: -8,5	M5: -5,5
Mucor ambiguus (NBRC 6742, (+)-Kreuzungstyp)						
Seq1	GAN04332	NCBI	verengt	offen	M2: -7,5	M9: -5,5
Seq2	GAN08850	NCBI	verengt ~	verengt ~	M1: -8,3	M9: -7,7
Seq3	GAN01222	NCBI	verengt	offen	M1: -10,3	M8: -6,1
Seq4	GAN05122	NCBI	verengt	offen	M2: -9,1	M5: -6,0
Seq5	GAN04526	NCBI	offen	offen	M1: -11,1	M10: -5,4

Tabelle 3.10: (Fortsetzung)

Nr.	Protein-ID	Datenbank	Enzymaktivität (Zugänglichkeit des aktiven Zentrums)		AutoDock Vina Freie Bindungsenergie (kcal/mol)	
			NADP	Trisporoid	NADP	Trisporoid
Mucor circinelloides (CBS 277.49, (–)-Kreuzungstyp)						
Seq1	150877	JGI / MycoCosm	verengt	offen	M1: -10,1	M8: -7,1
Seq2	155290	JGI / MycoCosm	verengt	verengt	M1: -10,1	M8: -4,8
Seq3	155026	JGI / MycoCosm	verengt	verengt	M4: -7,9	M1: -5,8
Seq4	45216	JGI / MycoCosm	offen	offen	M3: -9,2	M5: -5,6
Seq5	150779	JGI / MycoCosm	offen	offen	M1: -10,4	M8: -6,0
Seq6	139712	JGI / MycoCosm	verengt	offen	M4: -9,5	M2: -6,1
Mucor circinelloides (1006PhL, (+)-Kreuzungstyp)						
Seq1	EPB81768	NCBI	verengt	offen	M3: -9,3	M10: -7,5
Seq2	EPB87560	NCBI	verengt	verengt	M1: -9,7	M10: -6,6
Seq3	EPB84490	NCBI	verengt	verengt	M7: -8,0	M1: -8,3
Seq4	EPB84764	NCBI	offen	offen	M1: -9,2	M9: -6,3
Seq5	EPB81770	NCBI	offen	geschlossen	M1: -9,7	M1: -6,5
Seq6	EPB81046	NCBI	offen	offen	M7: -9,8	M6: -6,7
Mucor heterogamus (NRRL 1489, homothallisch)						
Seq1	524745	JGI / MycoCosm	verengt	verengt	M1: -10,4	M1: -9,3
Seq2	490332	JGI / MycoCosm	verengt	offen	M1: -7,5	M1: -7,7
Mucor mucedo (CBS 109.16, (–)-Kreuzungstyp)						
Seq1	<u>CAA98021</u>	NCBI	verengt	offen	M1: -8,7	M1: -6,8
Mycotypha africana (NRRL 2978, homothallisch)						
Seq1	816372	JGI / MycoCosm	verengt	offen	M1: -9,8	M6: -6,5
Seq2	875367	JGI / MycoCosm	verengt	offen	M1: -8,8	M5: -6,7
Seq3	812904	JGI / MycoCosm	offen	offen	M4: -9,6	M5: -6,4
Parasitella parasitica (CBS 412.66 NGI-315 ade-, (+)-Kreuzungstyp)						
Seq1	<u>CEP13671</u>	NCBI	verengt	offen	M1: -10,2	M5: -6,5
Seq2	CEP14056	NCBI	verengt	geschlossen	M1: -7,8	M1: -5,6
Seq3	CEP10423	NCBI	verengt	offen	M3: -9,5	M7: -6,2
Seq4	CEP10425	NCBI	offen	verengt	M3: -8,8	M8: -6,0
Seq5	CEP13270	NCBI	offen	verengt	M8: -9,5	M5: -6,4
Seq6	CEP18248	NCBI	offen	offen	M1: -8,7	M3: -7,8
Parasitella parasitica (NRRL 2501, (–)-Kreuzungstyp)						
Seq1	394004	JGI / MycoCosm	offen	offen	M1: -9,9	M6: -6,3
Seq2	368188	JGI / MycoCosm	verengt	offen	M2: -9,4	M6: -6,8
Seq3	446323	JGI / MycoCosm	offen	offen	M3: -9,1	M8: -5,8
Seq4	364222	JGI / MycoCosm	geschlossen	offen	M4: -6,8	M2: -7,5
Seq5	358854	JGI / MycoCosm	offen	offen	M1: -9,1	M8: -6,5

Tabelle 3.10: (Fortsetzung)

Nr.	Protein-ID	Datenbank	Enzymaktivität (Zugänglichkeit des aktiven Zentrums)		AutoDock Vina Freie Bindungsenergie (kcal/mol)	
			NADP	Trisporoid	NADP	Trisporoid
Phycomyces blakesleeanus (NRRL 1555, (+)-Kreuzungstyp)						
Seq1	124728	JGI / MycoCosm	offen ~	offen ~	M6: -8,0	M5: -6,1
Rhizomucor pusillus (NBRC 4578, (+)-Kreuzungstyp)						
Seq1	BAN58735	NCBI	verengt	verengt	M8: -9,3	M9: -6,5
Seq2	BAM29042	NCBI	verengt	verengt	M1: -9,8	M3: -6,2
Rhizopus microsporus (ATCC 52813, (+)-Kreuzungstyp)						
Seq1	248410	JGI / MycoCosm	offen	verengt	M4: -8,8	M5: -6,9
Seq2	206793	JGI / MycoCosm	verengt	offen	M3: -8,0	M3: -7,9
Rhizopus oryzae (RA 99-880, (+)-Kreuzungstyp)						
Seq1	740	JGI / MycoCosm	offen	offen	M1: -9,1	M4: -5,5
Syncephalastrum racemosum (NRRL 2496, homothallisch)						
Seq1	471093	JGI / MycoCosm	offen ~	offen ~	M2: -7,1	M7: -5,3
Seq2	477980	JGI / MycoCosm	offen	offen	M4: -8,8	M3: -7,2
Seq3	490367	JGI / MycoCosm	verengt	offen	M9: -7,9	M1: -9,3
Seq4	499703	JGI / MycoCosm	verengt	offen	M5: -6,6	M3: -7,6
Seq5	556430	JGI / MycoCosm	offen	offen	M7: -8,6	M1: -7,1
Seq6	522060	JGI / MycoCosm	verengt	offen	M1: -8,7	M5: -6,6
Thamnidium elegans (NRRL 2467, (+)-Kreuzungstyp)						
Seq1	407412	JGI / MycoCosm	verengt	verengt	M1: -6,3	M7: -5,2
Umbelopsis ramanniana (AG, (-)-Kreuzungstyp)						
Seq1	241297	JGI / MycoCosm	verengt ~	offen ~	M1: -8,2	M8: -6,5

Tabelle 3.10: (Fortsetzung)

3.3.2 Substratspezifität für Trisporoide der B- und E-Serie in TSP1

Entscheidend für die Trisporoid-Bindung ist die Spezies-abhängige Präferenz für bestimmte Trisporoid-Arten. Die Struktur innerhalb der Familie der *Mucorales* wurde mit Hilfe von Aktin, ITS-Regionen (internal transcribed spacer), Translationselongationsfaktor 1-alpha, sowie der großen und kleinen Untereinheiten ribosomaler DNA von Hoffmann et al. (2013) untersucht. Daraus ergab sich ein phylogenetischer Baum, in dem sich die von Protein-Ligand-Docking bestätigte Substratpräferenz der einzelnen Pilze widerspiegelt. Damit ist vor allem die Präferenz für Trisporoide der B-Serie in *Mucor mucedo* und für Trisporoide der E-Serie in *Phycomyces blakesleeanus* (Schimek und Wöstemeyer, 2009; Sutter et al., 1989) gemeint. Ein Trisporoid-Ligand hat verschiedene Möglichkeiten, mit dem Cosubstrat NADP so in Kontakt zu kommen, das eine Reaktion ermöglicht wird. Diese Molekülkonformationen unterscheiden sich zwischen den Trisporoid-Serien und verschiedenen TSP1-Proteinen.

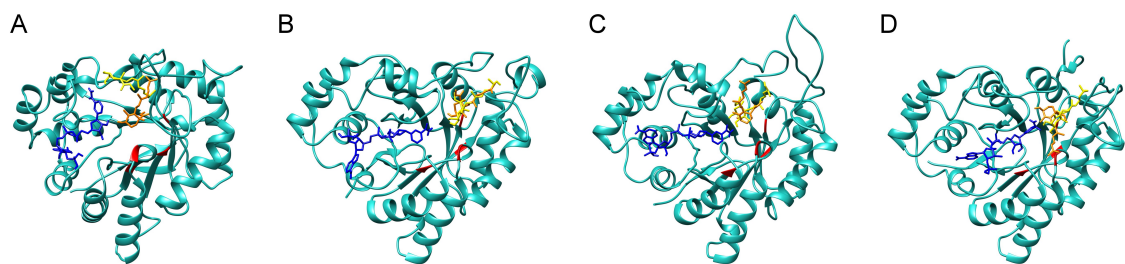


Abbildung 3.34: Unterschiede bei der Bindung von Trisporoiden der B-Serie (orange) und E-Serie (gelb) in TSP1-Proteinen aus (A) *Phycomyces blakesleeanus* (JGI: 124728), (B) *Mucor mucedo* (NCBI: CAA98021), (C) *Parasitella parasitica* (NCBI: CEP13671) und (D) *Absidia glauca* (NCBI: SAM02260). Das aktive Zentrum der Proteine ist rot und das Cosubstrat NADP blau markiert.

Abbildung 3.34 zeigt die unterschiedliche Positionierung von 4-Dihydromethyltrisporat der B- und E-Serie (TB1 und TE1) in TSP1-Proteinen aus verschiedenen Zygomyzeten. Bei *Absidia glauca* und *Phycomyces blakesleeanus* liegen TB1 und TE1 am deutlichsten auseinander. Bei *Absidia glauca* unterscheidet sich die Lage der Trisporoid-Seitenketten, während sich die Ringe an einer ähnlichen Position befinden. Bei *Phycomyces blakesleeanus* ist es genau umgekehrt. Der Endpunkt der Trisporoid-Seitenketten ist gleich und die Ringe liegen an unterschiedlichen Positionen. Der Ring, an dem die Reaktion stattfindet, liegt bei dem Trisporoid der E-Serie näher an dem Cosubstrat NADP. Eine Reaktion mit dem Trisporoid der B-Serie in der dargestellten Position ist unwahrscheinlich, da der Abstand zwischen Substrat und Cosubstrat zu groß ist.

Die freie Energie der Trisporoide unterscheidet sich in den einzelnen TSP1-Proteinen. Bei *Absidia glauca* (TB1: -6,7 kcal/mol, TE1: -6,1 kcal/mol) und *Mucor mucedo* (TB1: -6,8 kcal/mol, TE1: -5,8 kcal/mol) ist die Energie der Trisporoide der B-Serie niedriger. Bei *Parasitella parasitica* (TB1: -6,5 kcal/mol, TE1: -6,3 kcal/mol) ist ebenfalls die Ener-

gie der Trisporoide der B-Serie niedriger. Allerdings ist hier der Unterschied zwischen den beiden Trisporoid-Serien geringer. Bei *Phycomyces blakesleeen* (TB1: -6,1 kcal/mol, TE1: -6,5 kcal/mol) hingegen ist die Energie des Trisporoids der E-Serie niedriger. Auch hier ist der Unterschied zwischen den beiden Trisporoid-Serien eher gering.

3.3.3 Ligandenbindung bei TSP1 in *Parasitella parasitica*

Zygomyceten besitzen oft mehr als ein TSP1. Die verschiedenen TSP1-Isomere unterscheiden sich in ihren Eigenschaften in Bezug auf die Proteinoberfläche und Bindetaschen. Abbildung 3.35 zeigt die unterschiedlichen Konformationen von Substrat und Co-substrat nach Protein-Ligand-Docking mit 4-Dihydromethyltrisporat der B-Serie in den sechs TSP1-Varianten aus *Parasitella parasitica*. Die Ringe von NADP und Trisporoid liegen immer in vergleichbarer Richtung zueinander, während besonders die Seitenkette des Trisporoids beliebige Positionen einnehmen kann. Das Trisporoid unten rechts in der Abbildung gehört zu einem inaktiven TSP1-Protein (NCBI: CEP14056) mit geschlossener Bindetasche. Das Trisporoid liegt auf der Oberfläche des Enzyms.

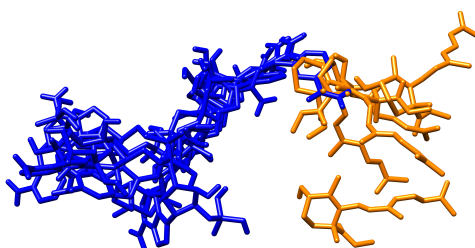


Abbildung 3.35: Unterschiede bei der Bindung von Trisporoiden der B-Serie (orange) in den verschiedenen TSP1-Isomeren aus *Parasitella parasitica*. Die für die sechs TSP1-Proteine mit AutoDock Vina ermittelten Ligandenkonformationen des Trisporoids (orange) und des Cosubstrats NADP (blau) wurden übereinandergelegt, um die Variabilität der Ligandenbindung zu veranschaulichen.

3.3.4 Ligandenbindung bei TSP2

Die Zugänglichkeit des aktiven Zentrums variiert in den einzelnen TSP2-Isomeren nicht so stark wie in TSP1-Isomeren. Die Bindetasche für das Cosubstrat NADP ist verzweigt und oft verengt. Die Bindetasche für das Trisporoid ist meist ein enger Tunnel. Die Ergebnisse des Protein-Ligand-Dockings zeigen, dass diese jedoch in den meisten Fällen noch groß genug sind, um NADP und Trisporoid so nah aneinander herankommen zu lassen, dass eine Reaktion möglich ist. Tabelle 3.11 zeigt, die Ergebnisse des Protein-Ligand-Dockings für die betrachteten TSP2-Proteine und die Zugänglichkeit des aktiven Zentrums der Enzyme für Substrat und Cosubstrat.

Unter den betrachteten TSP2-Proteinen gibt es nur eines, das keine offen zugängliche Bindetasche für Trisporoide besitzt (JGI: 6684 aus *Rhizopus oryzae*).

TSP2-Proteine, bei denen die Bindetaschen für Substrat und Cosubstrat miteinander verbunden sind, treten dagegen häufiger auf (NCBI: SAL99460 aus *Absidia glauca*, JGI: 437418 aus *Mucor heterogamus* und JGI: 826904 aus *Mycotypha africana*), wenn auch nicht so oft, wie bei TSP1-Proteinen.

Nr.	Protein-ID	Datenbank	Enzymaktivität (Zugänglichkeit des aktiven Zentrums)		AutoDock Vina Freie Bindungsenergie (kcal/mol)	
			NADP	Trisporoid	NADP	Trisporoid
Absidia glauca (CBS 101.48 RVII-324 met-, (-)-Kreuzungstyp)						
Seq1	SAL99460	NCBI	offen ~	offen ~	M3: -7,5	M2: -6,2
Seq2	SAL96112	NCBI	verengt	verengt	M2: -8,0	M4: -7,1
Absidia repens (NRRL 1336, (+)-Kreuzungstyp)						
Kein TSP2 identifiziert						
Backusella circina (FSU 941, (+)-Kreuzungstyp)						
Seq1	326016	JGI / MycoCosm	offen	offen	M5: -8,1	M5: -6,6
Seq2	248409	JGI / MycoCosm	verengt	verengt	M6: -7,6	M2: -7,5
Seq3	290299	JGI / MycoCosm	offen	verengt	M2: -9,3	M3: -6,7
Blakeslea trispora (NRRL 2456, (+)-Kreuzungstyp)						
Seq1	407222	JGI / MycoCosm	offen	offen	M4: -8,5	M9: -5,8
Seq2	340806	JGI / MycoCosm	offen	offen	M4: -8,0	M2: -5,4
Seq3	423722	JGI / MycoCosm	offen	offen	M1: -9,0	M8: -5,9
Seq4	434886	JGI / MycoCosm	offen	offen	M1: -8,7	M1: -5,9
Seq5	451430	JGI / MycoCosm	verengt	offen	M2: -9,9	M3: -6,7
Seq6	451925	JGI / MycoCosm	verengt	verengt	M7: -6,2	M1: -7,2
Hesseltinella vesiculosa (NRRL3301, (-)-Kreuzungstyp)						
Seq1	1338919	JGI / MycoCosm	verengt	verengt	M3: -10,1	M1: -6,4

Tabelle 3.11: Ergebnisse des Protein-Ligand-Dockings mit TSP2-Sequenzen aus Zygomyceten. Aufgelistet sind die besten Modelle für die Bindung des Cosubstrats NADP und eines Trisporoid-Liganden. Die Ziffer hinter dem M bezeichnet die Position des Modells in der Rangfolge der Ausgabe.

Unterstrichene Sequenzen wurden experimentell als TSP2 identifiziert.

~: Die Bindetaschen für NADP und Trisporoid liegen offen zugänglich an der Oberfläche des Proteins und sind nicht voneinander getrennt, wie bei der Mehrheit der TSP2-Proteine.

Nr.	Protein-ID	Datenbank	Enzymaktivität (Zugänglichkeit des aktiven Zentrums)		AutoDock Vina Freie Bindungsenergie (kcal/mol)	
			NADP	Trisporoid	NADP	Trisporoid
Lichtheimia corymbifera (FSU 9682, (–)-Kreuzungstyp)						
Seq1	CDH55711	NCBI	verengt	verengt	M9: -8,0	M3: -6,8
Lichtheimia hyalospora (FSU 10163, (–)-Kreuzungstyp)						
Seq1	34841	JGI / MycoCosm	verengt	verengt	M2: -8,8	M5: -6,3
Lichtheimia ramosa (FSU 6197, (–)-Kreuzungstyp)						
Seq1	CDS12894	NCBI	verengt	verengt	M2: -10,3	M1: -6,6
Mucor ambiguus (NBRC 6742, (+)-Kreuzungstyp)						
Seq1	GAN05628	NCBI	offen	offen	M1: -9,8	M10: -5,6
Seq2	GAN05629	NCBI	verengt	verengt	M3: -8,6	M4: -7,3
Seq3	GAN05647	NCBI	verengt	verengt	M2: -8,8	M1: -6,9
Seq4	GAN05649	NCBI	verengt	offen	M7: -8,2	M1: -6,8
Mucor circinelloides (CBS 277.49, (–)-Kreuzungstyp)						
Seq1	79258	JGI / MycoCosm	verengt	offen	M1: -10,0	M6: -6,8
Seq2	105700	JGI / MycoCosm	verengt	offen	M3: -9,0	M3: -6,2
Seq3	107961	JGI / MycoCosm	verengt	offen	M2: -9,0	M8: -5,6
Seq4	141495	JGI / MycoCosm	offen	offen	M1: -9,7	M1: -7,1
Seq5	142722	JGI / MycoCosm	verengt	offen	M3: -7,2	M2: -6,4
Seq6	170452	JGI / MycoCosm	verengt	verengt	M4: -8,5	M3: -6,9
Mucor circinelloides (1006PhL, (+)-Kreuzungstyp)						
Seq1	EPB87850	NCBI	offen	offen	M1: -9,6	M5: -6,1
Seq2	EPB82491	NCBI	verengt	verengt	M2: -9,6	M10: -3,6
Seq3	EPB87851	NCBI	verengt	verengt	M1: -9,2	M2: -6,9
Seq4	EPB87863	NCBI	verengt	offen	M6: -8,7	M5: -6,3
Seq5	EPB89015	NCBI	offen	offen	M3: -7,9	M9: -5,7
Mucor heterogamus (NRRL 1489, homothallisch)						
Seq1	461836	JGI / MycoCosm	offen	verengt	M2: -7,1	M4: -6,5
Seq2	437417	JGI / MycoCosm	offen	verengt	M1: -8,0	M4: -7,2
Seq3	525771	JGI / MycoCosm	offen	offen	M1: -9,6	M2: -6,7
Seq4	437418	JGI / MycoCosm	offen ~	offen ~	M3: -8,8	M5: -5,3
Seq5	396586	JGI / MycoCosm	verengt	offen	M1: 8,5	M2: -6,6
Mucor mucedo (CBS109.16, (–)-Kreuzungstyp)						
Seq1	<u>CAP72566</u>	NCBI	verengt	verengt	M3: -8,0	M4: -7,4
Mycotypha africana (NRRL 2978, homothallisch)						
Seq1	445142	JGI / MycoCosm	verengt	offen	M9: -7,7	M9: -5,8
Seq2	826904	JGI / MycoCosm	verengt ~	verengt ~	M8: -8,0	M3: -6,6

Tabelle 3.11: (Fortsetzung)

Nr.	Protein-ID	Datenbank	Enzymaktivität (Zugänglichkeit des aktiven Zentrums)		AutoDock Vina Freie Bindungsenergie (kcal/mol)	
			NADP	Trisporoid	NADP	Trisporoid
Parasitella parasitica (CBS 412.66 NGI-315 ade-, (+)-Kreuzungstyp)						
Seq1	CEP13616	NCBI	offen	offen	M6: -8,2	M4: -6,3
Seq2	CEP13991	NCBI	verengt	verengt	M9: -7,6	M3: -6,8
Parasitella parasitica (NRRL 2501, (-)-Kreuzungstyp)						
Seq1	385907	JGI / MycoCosm	offen	verengt	M1: -8,8	M2: -7,8
Seq2	424175	JGI / MycoCosm	offen	offen	M3: -6,9	M7: -6,4
Phycomyces blakesleeanus (NRRL1555, (+)-Kreuzungstyp)						
Seq1	124486	JGI / MycoCosm	offen	offen	M1: -8,7	M5: -5,7
Seq2	21973	JGI / MycoCosm	verengt	verengt	M6: -7,2	M5: -5,4
Seq3	118385	JGI / MycoCosm	offen	offen	M1: -8,6	M4: -6,9
Seq4	168867	JGI / MycoCosm	verengt	offen	M1: -9,2	M6: -5,9
Rhizopus microsporus (ATCC 52813, (+)-Kreuzungstyp)						
Seq1	259405	JGI / MycoCosm	verengt	offen	M10: -6,1	M8: -5,5
Seq2	217377	JGI / MycoCosm	offen	verengt	M1: -8,1	M5: -4,6
Rhizopus oryzae (RA 99-880, (+)-Kreuzungstyp)						
Seq1	6648	JGI / MycoCosm	offen	verengt	M1: -9,8	M5: -6,8
Seq2	6684	JGI / MycoCosm	verengt	geschlossen	M1: -8,2	M1: -4,5
Seq3	6689, 6691	JGI / MycoCosm	offen	verengt	M1: -9,3	M2: -7,1
Seq4	12864	JGI / MycoCosm	verengt	offen	M6: -8,3	M7: -6,2
Syncephalastrum racemosum (NRRL 2496, homothallisch)						
Seq1	546915	JGI / MycoCosm	verengt	offen	M1: -9,3	M2: -6,7
Thamnidium elegans (NRRL 2467, (+)-Kreuzungstyp)						
Seq1	430843	JGI / MycoCosm	verengt	offen	M9: -8,1	M2: -6,6
Seq2	369045	JGI / MycoCosm	verengt	verengt	M6: -7,7	M6: -6,0
Seq3	391137	JGI / MycoCosm	verengt	offen	M1: -9,6	M5: -6,5
Umbelopsis ramanniana (AG, (-)-Kreuzungstyp)						
Seq1	232239	JGI / MycoCosm	verengt	offen	M1: -8,1	M3: -6,5
Seq2	232236	JGI / MycoCosm	verengt	verengt	M1: -8,5	M1: -7,1

Tabelle 3.11: (Fortsetzung)

3.3.5 Substratspezifität für Trisporoide der B- und E-Serie in TSP2

Abbildung 3.36 zeigt die unterschiedliche Positionierung von 4-Dihydrotrisporin der B- und E-Serie (TB2 und TE2) in TSP2-Proteinen aus verschiedenen Zygomyceten. Bei *Phycomyces blakesleeanus* liegen TB2 und TE2 deutlich auseinander. Bei *Mucor mucedo*, *Parasitella parasitica* und *Absidia glauca* liegen die beiden Trisporoide an vergleichbaren Positionen.

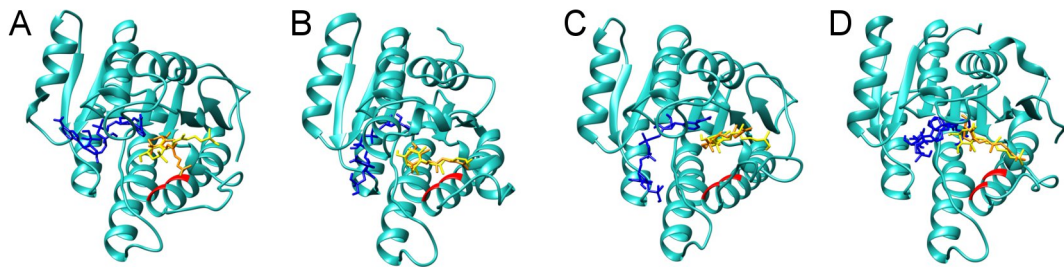


Abbildung 3.36: Unterschiede bei der Bindung von Trisporoiden der B-Serie (orange) und E-Serie (gelb) in TSP2-Proteinen aus (A) *Phycomyces blakesleeanus* (JGI: 124486), (B) *Mucor mucedo* (NCBI: CAP72566), (C) *Parasitella parasitica* (NCBI: CEP13616) und (D) *Absidia glauca* (NCBI: SAL99460). Das aktive Zentrum der Proteine ist rot und das Cosubstrat NADP blau markiert.

Die freie Energie der Trisporoide unterscheidet sich in den einzelnen TSP2-Proteinen. Bei *Absidia glauca* (TB2: -6,2 kcal/mol, TE2: -5,3 kcal/mol) und *Mucor mucedo* (TB2: -7,4 kcal/mol, TE2: -7,0 kcal/mol) ist die Energie der Trisporoide der B-Serie niedriger. Bei *Parasitella parasitica* (TB2: -7,3 kcal/mol, TE2: -7,1 kcal/mol) ist ebenfalls die Energie des Trisporoids der B-Serie niedriger. Allerdings ist hier der Unterschied zwischen den beiden Trisporoid-Serien geringer. Bei *Phycomyces blakesleeanus* (TB2: -5,7 kcal/mol, TE2: -6,0 kcal/mol) hingegen ist die Energie des Trisporoids der E-Serie niedriger. Auch hier ist der Unterschied zwischen den beiden Trisporoid-Serien eher gering.

3.3.6 Ligandenbindung bei TSP2 in *Parasitella parasitica*

Zygomyceten besitzen bis zu sechs TSP2-Proteine. Die verschiedenen TSP2-Isomere unterscheiden sich in ihren Eigenschaften in Bezug auf die Proteinoberfläche und Binde-taschen. Abbildung 3.37 zeigt die unterschiedlichen Konformationen von Substrat und Cosubstrat nach Protein-Ligand-Docking mit 4-Dihydrotrisporin der B-Serie in den beiden TSP2-Varianten aus *Parasitella parasitica*.

Der Nicotinamid-Ring des Cosubstrats NADP befindet sich in beiden TSP2-Proteinen an einer vergleichbaren Position. Das Adenin des NADP ist entlang der Achse aus NADP und Trisporoid gespiegelt. Da die Bindetasche in einem der TSP2-Proteine verengt ist, liegt das Trisporoid weiter vom NADP entfernt, als im anderen TSP2-Protein. Die beiden Trisporoide liegen in einem Winkel von 90° zueinander.

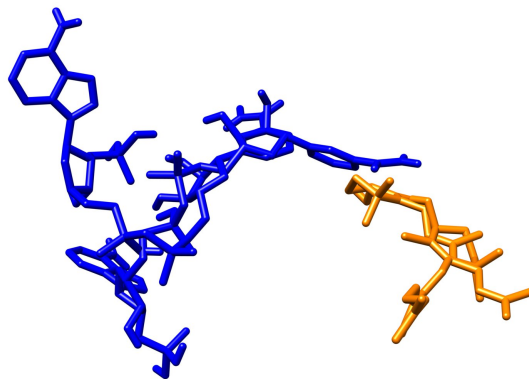


Abbildung 3.37: Unterschiede bei der Bindung von Trisporoiden der B-Serie (orange) in den beiden TSP2-Isomeren aus *Parasitella parasitica*. Die für die TSP2-Proteine mit AutoDock Vina ermittelten Ligandenkonformationen des Trisporoids (orange) und des Cosubstrats NADP (blau) wurden übereinandergelegt, um die Variabilität der Ligandenbindung zu veranschaulichen.

3.3.7 Gemeinsamkeiten in der Ligandenbindung bei TSP1 und TSP2

Die beiden Dehydrogenasen TSP1 und TSP2 sind NADP-abhängig und zeigen eine Oxioreduktaseaktivität für Substrate mit nah verwandter Struktur, Trisporoide. Obwohl sie ähnliche Reaktionen katalysieren, gehören sie zu unterschiedlichen Enzymklassen mit jeweils spezifischem Proteinaufbau. TSP1 ist eine Aldo-Keto-Reduktase und zeigt die typische TIM-Barrel-Struktur dieser Dehydrogenasen. TSP2 ist eine kurzkettige Dehydrogenase mit einer Rossmann-Faltung. Die aktiven Zentren von TSP1 (Abbildung 3.38, linke Seite) und TSP2 (Abbildung 3.38, rechte Seite) beinhalten beide ein Lysin und ein Tyrosin. Betrachtet man diese beiden Aminosäuren unabhängig von ihrer Anordnung innerhalb der vollständigen Proteinstrukturen, zeigt sich, dass ihre Orientierung in TSP1 und TSP2 sehr ähnlich ist (Abbildung 3.38, Mitte). Die Achsen, die aus dem Cosubstrat NADP und dem Trisporoid entstehen, liegen bei TSP1- und TSP2-Proteinen fast in einem Winkel von 90° zueinander.

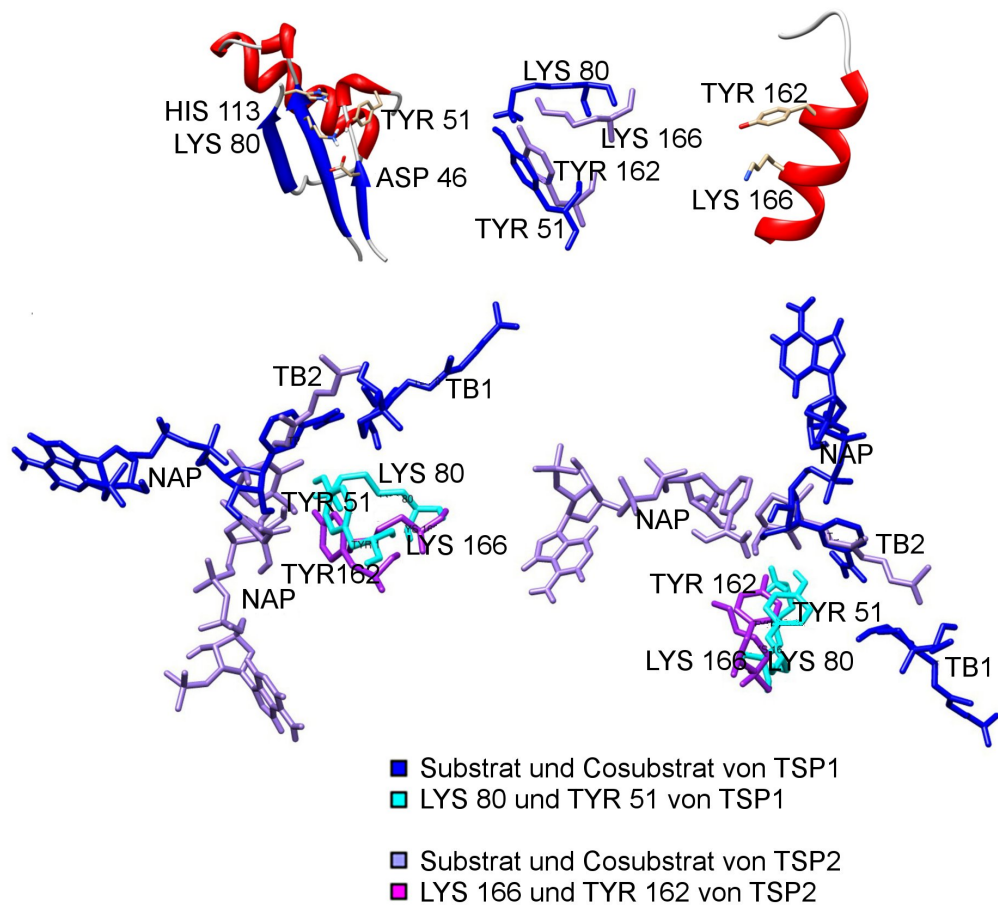


Abbildung 3.38: Unterschiede bei der Bindung von Trisporoiden an TSP1- und TSP2-Proteinen aus *Mucor mucedo*. Die aktiven Zentren von TSP1 (Aspartat, Tyrosin, Lysin, Histidin) und TSP2 (Tyrosin, Lysin) wurden so übereinandergelegt, dass die Lysine und Tyrosine eine ähnliche Position einnehmen. Die beiden Achsen, die sich aus den gebundenen Trisporoiden und Cosubstraten ergeben, sind in TSP1 und TSP2 um fast 90° zueinander verdreht.

3.4 Untersuchung der Bildung von Homo- und Heterodimeren

Die Vorlage für die Modellierung von TSP1-Proteinen, eine Xylose-Reduktase aus *Candida tenuis*, bildet Homodimere aus. Klimacek et al. (2003) haben gezeigt, dass einige Aminosäuren innerhalb der $\alpha 5$ -Helix entscheidend für die Dimerbildung sind. Ala-173 und Leu-174 wurden von zwei Argininen, und Arg-180 von einem Alanin ersetzt, um den Einfluss dieser Aminosäuren auf die Dimerisierung der Xylose-Reduktase zu beweisen. Das multiple Sequenzalignment der TSP1-Sequenzen und Xylose-Reduktasen aus verschiedenen Pilzen in Abbildung 3.39 zeigt, dass sich die Vorlage der Modellierung an Hand zweier Aminosäuren in der $\alpha 5$ -Helix von TSP1-Proteinen unterscheidet. In TSP1-Proteinen stehen ein Glutamin oder Methionin anstelle des Ala-173 und ein Serin oder Threonin an der Position des Arg-180. Dieser Sequenzabschnitt unterstützt die Annahme, dass es sich bei der Xylose-Reduktase aus *Rhizomucor pusillus* um ein TSP1-Protein handelt. Die Aminosäuren an den Positionen 173 – 182 dieser Sequenz sind identisch mit denen im korrespondierenden Abschnitt 174 – 183 in der TSP1-Sequenz aus *Phycomyces blakesleeanus*.

Absidia glauca SAL98187	VMALLDLFTY
Mucor heterogamus 490332	VQSIVDLLTY
Parasitella parasitica CBS 412.66 CEP13671	CQAILDLLSY
Phycomyces blakesleeanus 124728	VQLILDLLTY
Rhizomucor pusillus BAN58735	CQAILDMLTY
Rhizomucor pusillus BAM29042	VQLILDLLTY
Arthroderma otae XP_002849477	PQLIMDLLRY
Aspergillus nidulans XP_658027	AQLLLDLLRY
Candida tenuis O74237	GALLLDLLRG
Kluyveromyces lactis XP_454929	GALIQLLLRG
Saccharomyces cerevisiae NP_011972	GSLIQLLLRG
Scheffersomyces stipitis XP_001385181	GALLLDLLRG
Trichoderma reesei XP_006965651	AQSVYDALIY
	: * :

Abbildung 3.39: Multiples Sequenzalignment der $\alpha 5$ -Helix einzelner TSP1-Sequenzen und verschiedener Xylose-Reduktasen. Die beiden Enzyme lassen sich anhand der Aminosäuren an zwei Positionen in dieser Region unterscheiden (fettgedruckt).

TSP1-Proteine besitzen auf ihrer Oberfläche Finger-förmige Ausbildungen, die aus bestimmten Aminosäuren bestehen. Abbildung 3.40 zeigt diese Bereiche an einem TSP1-Homodimer aus *Mucor circinelloides*. Die Schnittstelle zwischen den beiden TSP1-Monomeren ist symmetrisch und besteht aus zwei Arginin- und zwei Glutaminresten aus beiden Partnern. Auf der Proteinoberfläche sind weitere ähnlich angeordnete Aminosäuren, Glutamine und Lysine, zu sehen. Dabei handelt es sich um zusätzliche Protein-Bindestellen zur Bildung von Heterodimeren. Ein möglicher Partner für Interaktionen dieser Art sind die verschiedenen TSP1-Isomere, die viele Zygomyceten besitzen.

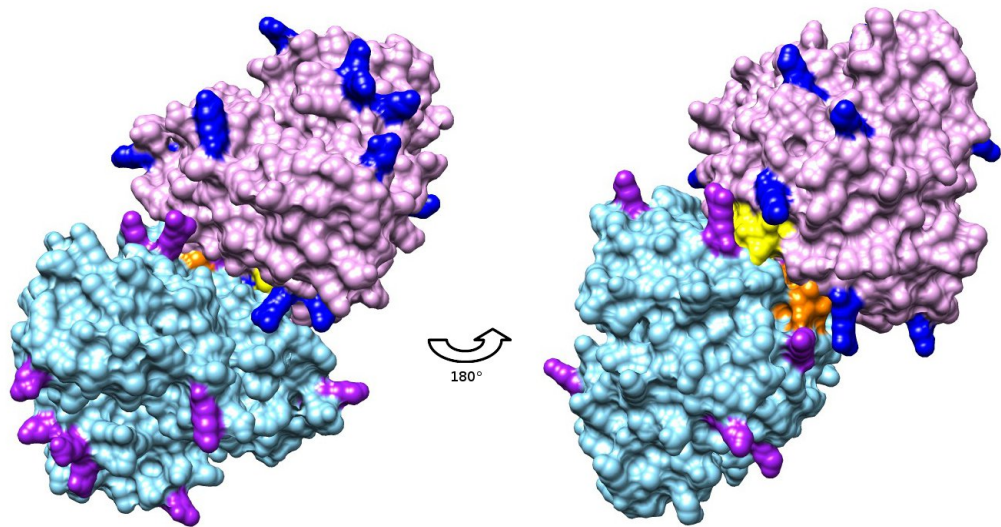


Abbildung 3.40: Modellierung eines TSP1-Homodimers für *Mucor circinelloides*. Die Schnittstelle zwischen den zwei TSP1-Untereinheiten ist gelb und orange markiert. Aminosäuren, die zusätzliche Protein-Bindestellen zur Bildung von Heterodimeren formen, sind blau und magenta markiert.

Ob auch TSP2 als Dimerisierungspartner dienen kann, wurde an Proteinen aus *Parasitella parasitica* getestet (Abschnitt 3.4.4).

3.4.1 Modellierung von TSP1-Dimeren in *Absidia glauca*

Absidia glauca besitzt vier TSP1-Proteine, die sich aufgrund ihrer C-terminalen Enden in zwei Gruppen einteilen lassen. Die beiden TSP1-Sequenzen vom Typ1 (SAL98947, SAM02260) zeigen die größte Sequenzähnlichkeit zu Original-TSP1-Proteinen und besitzen das C-terminale Sequenzmotiv G(L/F)DLPLFD. Die beiden TSP1-Sequenzen vom Typ2 (SAL98187, SAM06439) enthalten ein C-terminales Sequenzmotiv, das außer in *Absidia glauca* und *Absidia repens* auch in den drei betrachteten *Lichtheimia*-Stämmen, den beiden Stämmen von *Mucor circinelloides* und *Syncephalastrum racemosum* vorkommt. AQK[IL]P[IL]FY unterscheidet sich von anderen C-terminalen Motiven in TSP1-Sequenzen.

Abbildung 3.41 zeigt die modellierten Proteinstrukturen der in *Absidia glauca* vorkommenden TSP1-Varianten. Alle Proteine besitzen offene Bindetaschen für Substrat und Co-

substrat. Die Abbildung enthält die Ergebnisse aus dem Protein-Ligand-Docking. Die Lage der flexiblen Schleifenregionen unterscheidet sich zwischen den vier TSP1-Proteinen. Die flexible Schleifenregion der zweiten Struktur (NCBI: SAL98947) ragt deutlich aus dem Protein hervor und bietet sich als Inhibitor an (siehe Abschnitt 3.4.2).

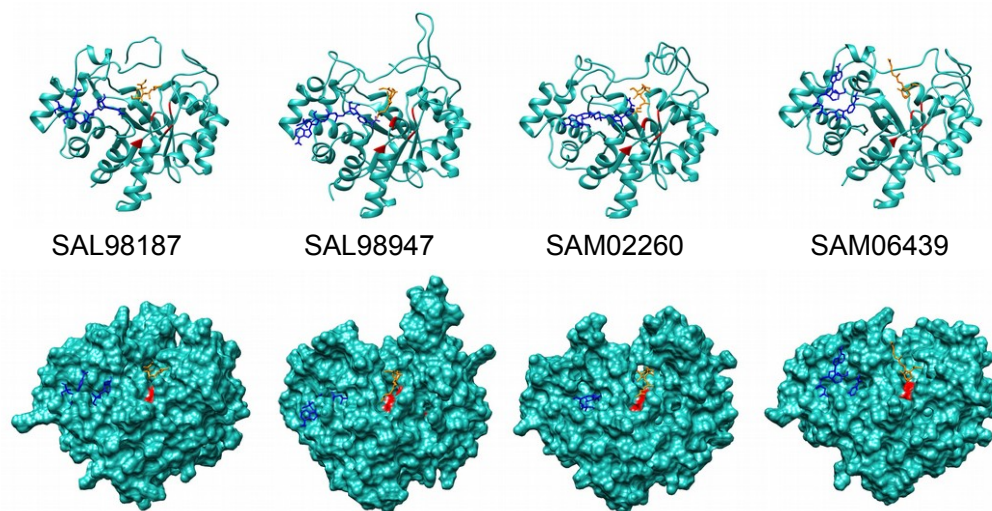


Abbildung 3.41: Modellierung der TSP1-Proteine aus *Absidia glauca*.

Die Schleifenregionen sind sehr variabel zwischen den verschiedenen TSP1-Isoformen. Zu erkennen sind Finger-förmige Ausbildungen auf der Oberfläche der Proteine, die eine Rolle bei der Dimerisierung spielen. Alle Proteinstrukturen beinhalten das beste Modell des Protein-Ligand-Dockings mit NADP (blau) und einem Trisporoid (orange). Das aktive Zentrum der Enzyme ist rot markiert.

Tabelle 3.12 zeigt die Ergebnisse aus dem Protein-Protein-Docking mit den vier TSP1-Proteinen aus *Absidia glauca*. Zu sehen ist die Aktivität der gebildeten Dimere. Die Aktivität eines Enzyms wird mit der Zugänglichkeit des aktiven Zentrums definiert.

	SAL98187	SAL98947	SAM02260	SAM06439
SAL98187	100 %	92 %	72 %	100 %
SAL98947	40 %	93 %	65 %	49 %
SAM02260	53 %	73 %	69 %	52 %
SAM06439	86 %	92 %	78 %	91 %

Tabelle 3.12: Prozent an enzymatisch aktiven TSP1-Dimeren aus *Absidia glauca*.

Aktivität wird über die Zugänglichkeit des aktiven Zentrums für Cosubstrat und Substrat definiert. Zeilen zeigen den Rezeptor, Spalten den Liganden des Protein-Protein-Dockings. Homodimere sind grau hinterlegt.

Die enzymatische Aktivität war bei den Dimeren aus TSP1-Proteinen aus *Absidia glauca* sehr stabil. Sie war unabhängig von den im Protein-Protein-Docking kombinierten TSP1-Monomeren mindestens zu 50 % gegeben. Dimere, die aus zwei TSP1-Proteinen vom selben Typ bestehen, waren aktiver, als solche, die sich aus einer Kombination zweier unterschiedlicher TSP1-Typen heraus bilden. Das bereits als TSP1 erkannte Original-

Protein war nach der Dimerisierung weniger aktiv, als die neu beschriebenen TSP1-Proteine vom zweiten Typ. Abbildung 3.42 zeigt einige der mit Protein-Protein-Docking erzeugten TSP1-Dimere aus *Absidia glauca*.

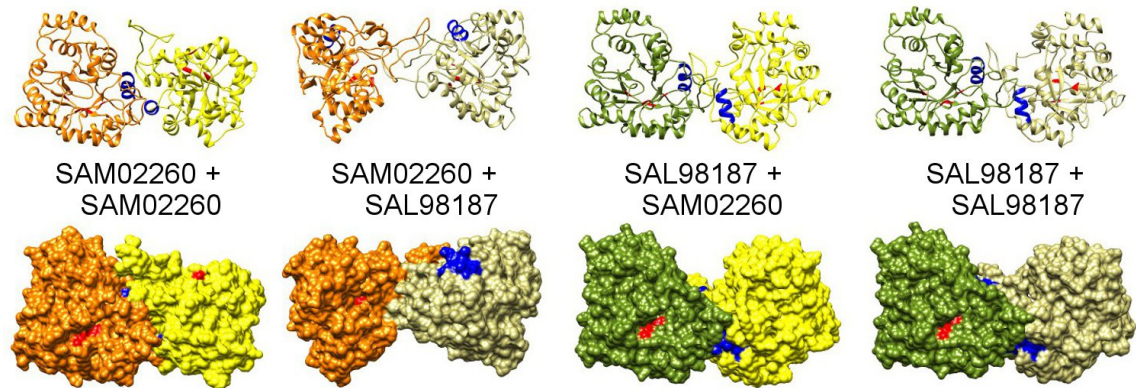


Abbildung 3.42: Modellierung der TSP1-Dimere aus den TSP1-Proteinen SAM02260 (Typ1) und SAL98187 (Typ2) aus *Absidia glauca*.

Die Dimere sind nicht symmetrisch, das heißt Rezeptorproteine (linker Partner) verhalten sich bei der Bindung anders als Ligandenproteine (rechter Partner). Die aktiven Zentren sind rot, die α 5-Helices blau markiert. Alle dargestellten Dimere sind zu mindestens 53 % enzymatisch aktiv.

3.4.2 Modellierung von TSP1-Dimeren in *Parasitella parasitica*

Parasitella parasitica besitzt sechs TSP1- und TSP1-ähnliche Proteine, die sich aufgrund ihrer C-terminalen Enden in vier Gruppen einteilen lassen. Die Original-TSP1-Sequenz vom Typ1 (CEP13671) zeigt die größte Sequenzähnlichkeit zu Original-TSP1-Proteinen und besitzt das C-terminale Sequenzmotiv GFGLPLFD. Die TSP1-Proteine vom Typ2 sind *Parasitella*-spezifische Sequenzen (CEP10423, CEP10425), die das C-terminale Sequenzmotiv GFDFP(V/L)FA besitzen. Die beiden TSP1-Proteine vom Typ3 (CEP14056, CEP18248) sind regulatorische Untereinheiten ohne enzymatische Aktivität, die für die posttranslationale Regulation von TSP1-Dimeren entscheidend sind. Sie besitzen das C-terminale Sequenzmotiv N(F/I)(N/R)(L/I)PLFN. Das TSP1-Protein vom Typ4 (CEP13270) ist ebenfalls ein regulatorisches Protein. Seine Aktivität unterscheidet, im Gegensatz zu den beiden TSP1-Proteinen vom Typ3, zwischen ursprünglichen und *Absidia*-spezifischen TSP1-Proteinen. Dieses TSP1-Protein war in Dimeren mit anderen TSP1-Proteinen aus *Parasitella parasitica* enzymatisch aktiv. Es besitzt das C-terminale Sequenzmotiv GFNLPLFN, das sich auch bei anderen Zygomyceten findet.

Abbildung 3.43 zeigt, wie unterschiedlich die modellierten Proteinstrukturen für TSP1 aus *Parasitella parasitica* an manchen Stellen sind. Die sechs TSP1-Proteine unterscheiden sich vor allem in der Schleifenregion (3.12, Region 2) und den Bindetaschen für NADP und Trisporoide. Alle Proteinmodelle haben offene Bindetaschen für Substrat und Cosubstrat, mit Ausnahme von CEP14056. Bei diesem Protein ist die Substratbindetasche

geschlossen. CEP13671 besitzt die flexibelste Schleifenregion. Bei CEP18248 fehlt das Lysin des aktiven Zentrums. Die anderen drei Aminosäuren des aktiven Zentrums liegen zu weit auseinander. Dieses Protein wird daher als enzymatisch inaktiv betrachtet. Es hat eine rein regulatorische Funktion.

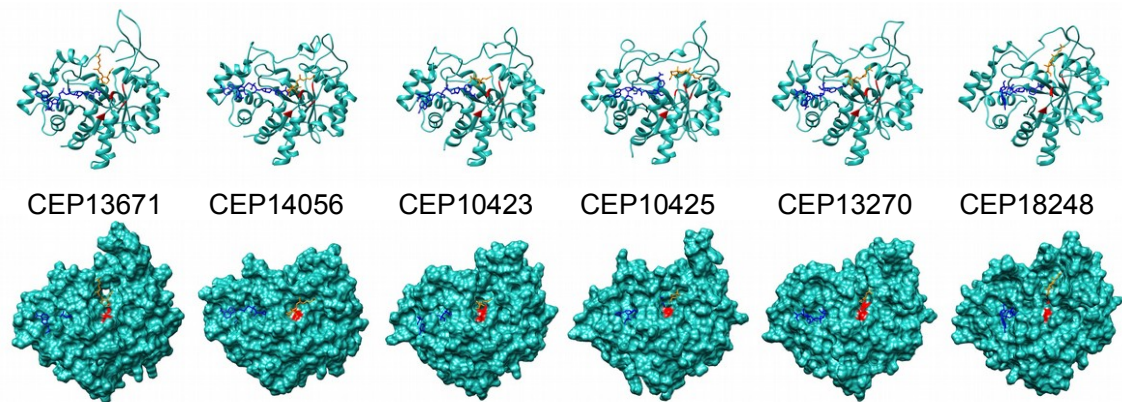


Abbildung 3.43: Modellierung der TSP1-Proteine aus *Parasitella parasitica*.

Die Schleifenregionen sind sehr variabel zwischen den verschiedenen TSP1-Isoformen. Zu erkennen sind Finger-förmige Ausbildungen auf der Oberfläche der Proteine, die eine Rolle bei der Dimerisierung spielen. Alle Proteinstrukturen beinhalten das beste Modell des Protein-Ligand-Dockings mit NADP (blau) und einem Trisporoid (orange). Das aktive Zentrum der Enzyme ist rot markiert.

Alle möglichen Kombinationen von TSP1-Varianten aus *Parasitella parasitica* wurden mit Hilfe von Protein-Protein-Docking simuliert und miteinander verglichen. Dabei ergab sich ein eindeutiges Muster in der theoretisch möglichen Aktivität der Enzyme. Tabelle 3.13 zeigt die Ergebnisse aus dem Protein-Protein-Docking mit den sechs TSP1-Proteinen aus *Parasitella parasitica*. Zu sehen ist die Aktivität der gebildeten Dimere. Die Aktivität eines Enzyms wird mit der Zugänglichkeit des aktiven Zentrums definiert.

	CEP13671	CEP14056	CEP10423	CEP10425	CEP13270	CEP18248
CEP13671	11 %	100 %	93 %	100 %	52 %	94 %
CEP14056	0 %	0 %	0 %	0 %	0 %	0 %
CEP10423	28 %	100 %	100 %	100 %	100 %	100 %
CEP10425	0 %	100 %	100 %	7 %	0 %	91 %
CEP13270	0 %	22 %	34 %	75 %	0 %	41 %
CEP18248	0 %	0 %	0 %	0 %	0 %	0 %

Tabelle 3.13: Prozent an enzymatisch aktiven TSP1-Dimeren aus *Parasitella parasitica*.

Aktivität wird über die Zugänglichkeit des aktiven Zentrums für Cosubstrat und Substrat definiert. Zeilen zeigen den Rezeptor, Spalten den Liganden des Protein-Protein-Dockings. Homodimere sind grau hinterlegt.

(Tabelle nach Ellenberger et al. (2017))

Die theoretische enzymatische Aktivität unterschied sich zwischen Homo- und Heterodimeren aus *Parasitella parasitica*. Homodimere bildeten meist Strukturen, bei denen

die Bindetaschen geschlossen waren oder vom Bindungspartner blockiert wurden. Ausschließlich Homodimere von CEP10423 waren in der Lage, ihre Bindetaschen offen zu halten, nachdem sie an einen Partner gebunden waren. Diese Eigenschaft blieb bei der Simulation der lokalen Proteinbewegung mit PELE erhalten. Heterodimere von CEP13671 zeigten eine größere Fähigkeit ihre Bindetaschen offen zu halten. In CEP13671 Homodimeren blockierte die flexible Schleifenregion das aktive Zentrum des Bindungspartners (Abbildung 3.44). Auf diese Weise wurde verhindert, dass Liganden Zugang zum aktiven Zentrum des Enzyms bekamen. Konsequenz daraus war die enzymatische Inaktivierung des Proteinkomplexes. Einige andere TSP1-TSP1-Kombinationen ermöglichten offene Bindetaschen und somit enzymatische Aktivität (Tabelle 3.13). Diese Bereiche waren zum Teil verengt und neigten dazu, sich in Simulationen der lokalen Proteinbewegung mit sinkendem Energielevel zu schließen. Dimere aus CEP14056 und CEP18248 zeigten keinerlei enzymatische Aktivität, wenn sie als Rezeptor im Protein-Protein-Docking definiert wurden. Als Ligand bildeten sie aktive Heterodimere mit CEP13671 und CEP10423. Daher sind diese beiden TSP1-Varianten als rein regulatorische Untereinheiten zu sehen.

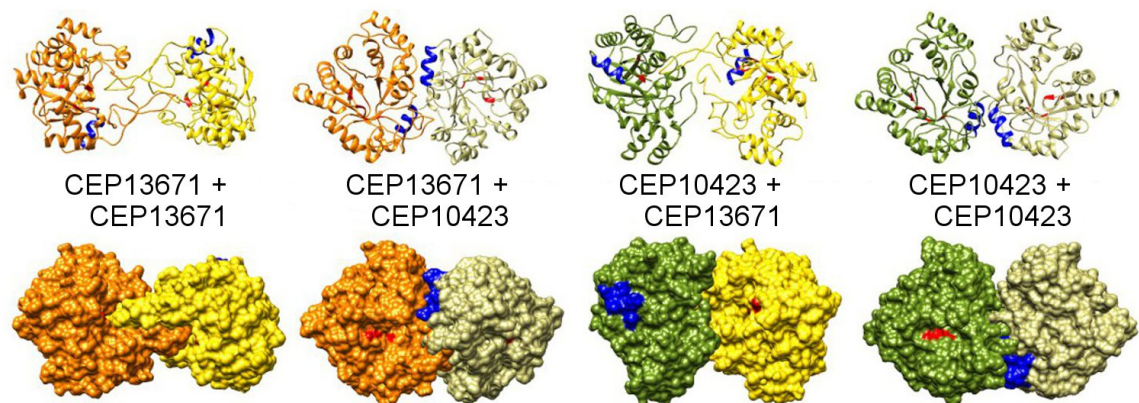


Abbildung 3.44: Modellierung der TSP1-Dimere aus den TSP1-Proteinen CEP13671 und CEP10423 aus *Parasitella parasitica*.

Die Dimere sind nicht symmetrisch, das heißt Rezeptorproteine (linker Partner) verhalten sich bei der Bindung anders als Ligandenproteine (rechter Partner). Die aktiven Zentren sind rot, die $\alpha 5$ -Helices blau markiert. Die aktiven Zentren des Homodimers aus zwei CEP13671 werden von den flexiblen Schleifenstrukturen der Partner blockiert und die Enzyme somit inaktiviert. (Abbildung nach Ellenberger et al. (2017))

Es gibt Unterschiede zwischen den Dimerisierungsbereichen von Xylose-Reduktasen und TSP1. Mit ClusPro wurden verschiedenen Proteinkomplexe modelliert. Die TSP1-Untereinheiten binden an unterschiedlichen Stellen. Der Kontakt über eine den Xylose-Reduktasen ähnliche Schnittstelle ist bei TSP1 möglich (Abbildung 3.44, Homodimer aus CEP10423). Zusätzlich gibt es mehrere weitere Konformationen. Dimere können zum Beispiel über die flexible Schleifenstruktur der TSP1-Proteine gebildet werden (Abbildung 3.44, Heterodimer aus CEP10423 und CEP13671). Dabei spielt die $\alpha 5$ -Helix nicht immer eine Rolle (Abbildung 3.44, Homodimer aus CEP13616, blaue Bereiche). Filme

der in Abbildung 3.44 dargestellten TSP1-Dimere finden sich auf dem beigefügten Datenträger. Video 1 zeigt die Inaktivierung von CEP13671 nach der Bildung eines Homodimers. Video 2 zeigt das aktive Homodimer von CEP10423. Video 3 demonstriert die Flexibilität der Schleifenregion zwischen dem β 4-Strang und der α 4-Helix der TSP1-Proteine. Das aktive Zentrum ist rot, die α 5-Helix blau und die flexible Schleifenregion grün markiert. Die Videos wurden der Publikation über posttranslationale Regulation innerhalb der Pheromonbiosynthese von *Parasitella parasitica* im *Journal of Theoretical Biology* als zusätzliches Material hinzugefügt (Ellenberger et al., 2017).

Die Energielevel der Monomere und Dimere bei der Untersuchung von Trajektorien lokaler Proteinbewegung wurden mit PELE berechnet und belegen die Existenz von Dimeren in *Parasitella parasitica*. Fünf von sechs Monomeren resultierten in positiven Energieleveln zu Beginn der Simulation. Nur das Proteinmodell der Sequenz CEP13270 ergab ein negatives Energielevel. Die Minima der PELE-Trajektorien waren für alle sechs TSP1-Proteine vergleichbar und lagen zwischen -12 444 kcal/mol und -12 262 kcal/mol. Die Energielevel der Dimere waren immer negativ mit deutlichen Unterschieden zwischen den Startkonformationen. Die Werte lagen hier zwischen -15 412 kcal/mol und -8 297 kcal/mol. Die Energieminima der Trajektorien waren zwischen den Dimeren vergleichbar und lagen zwischen -24 867 kcal/mol und -24 482 kcal/mol.

3.4.3 Modellierung von TSP2-Dimeren in *Parasitella parasitica*

Parasitella parasitica besitzt zwei Gene für TSP2-Proteine. Die möglichen Kombinationen der beiden TSP2-Varianten aus *Parasitella parasitica* wurden mit Hilfe von Protein-Protein-Docking simuliert und miteinander verglichen. Tabelle 3.14 zeigt die Ergebnisse aus dem Protein-Protein-Docking mit den beiden TSP2-Proteinen aus *Parasitella parasitica*. Zu sehen ist die Aktivität der gebildeten Dimere. Die Aktivität eines Enzyms wird mit der Zugänglichkeit des aktiven Zentrums definiert. Dimerisierung hatte einen negativen Einfluss auf die enzymatische Aktivität der TSP2-Proteine. Die beiden TSP2-Varianten unterschieden sich in ihrer enzymatischen Aktivität. Dimere mit dem TSP2-Protein CEP13991 als Ligand waren immer inaktiv. Dimere mit dem zweiten TSP2-Protein CEP13616 als Ligand besaßen zumindest eine geringe enzymatische Aktivität von circa 30 %.

	CEP13616	CEP13991
CEP13616	32 %	0 %
CEP13991	29 %	0 %

Tabelle 3.14: Prozent an enzymatisch aktiven TSP2-Dimeren aus *Parasitella parasitica*.

Aktivität wird über die Zugänglichkeit des aktiven Zentrums für Cosubstrat und Substrat definiert. Zeilen zeigen den Rezeptor, Spalten den Liganden des Protein-Protein-Dockings. Homodimere sind grau hinterlegt.

Abbildung 3.45 zeigt die modellierten TSP2-Dimere aus *Parasitella parasitica*. Die aktiven Zentren der TSP2-Proteine wurden nicht wie bei TSP1-Dimeren von der flexiblen Schleifenregion des Partnerproteins verschlossen. Bei den TSP2-Proteinen wurde die gesamte Struktur nach der Anlagerung des Partners zusammengestaucht, bis sich der Zugang zum aktiven Zentrum geschlossen hatte und das Enzym dadurch inaktiv wurde.

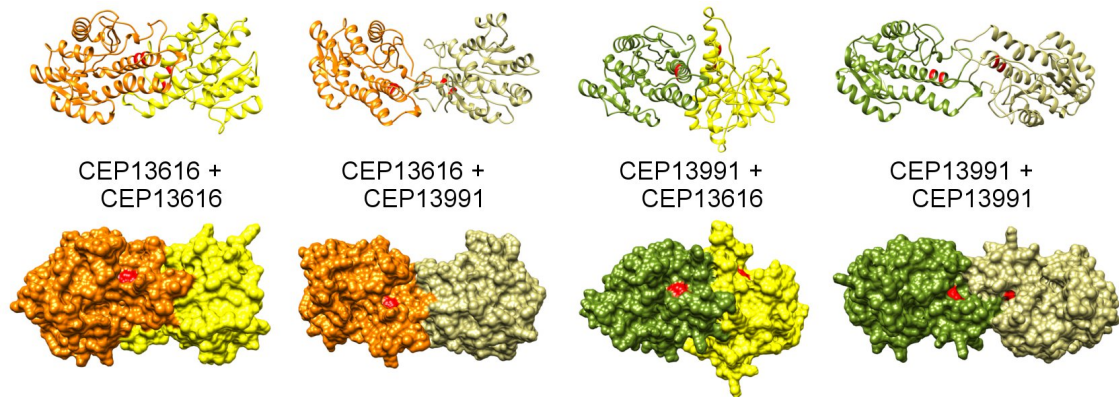


Abbildung 3.45: Modellierung der TSP2-Dimere aus *Parasitella parasitica*.

Die Dimere sind nicht symmetrisch, das heißt Rezeptorproteine (linker Partner) verhalten sich bei der Bindung anders als Ligandenproteine (rechter Partner). Das aktive Zentrum der Enzyme ist rot markiert.

3.4.4 Modellierung von TSP1-TSP2-Dimeren in *Parasitella parasitica*

In *Parasitella parasitica* wird die enzymatische Aktivität von TSP1-Proteinen posttranslational reguliert. Neben den verschiedenen TSP1-Varianten können weitere Proteine regulatorisch auf TSP1 einwirken. Die möglichen Kombinationen zwischen TSP1- und TSP2-Proteinen wurden mit Hilfe von Protein-Protein-Docking simuliert und miteinander verglichen. Tabelle 3.15 zeigt die Ergebnisse aus dem ersten Protein-Protein-Docking mit TSP1 als Rezeptor und TSP2 als Ligand. Zu sehen ist die Aktivität der gebildeten Dimere. Die Aktivität eines Enzyms wird mit der Zugänglichkeit des aktiven Zentrums definiert. CEP14056 und CEP18248 vom Typ2 hatten auch in Kombination mit TSP2 keine enzymatische Aktivität. Die TSP1-Proteine vom Typ2 und Typ4 (CEP10423, CEP10425, CEP13270) waren aktiver, als das Original-TSP1-Protein (CEP13671).

	CEP13616	CEP13991
CEP13671	79 %	62 %
CEP14056	0 %	0 %
CEP10423	100 %	75 %
CEP10425	100 %	85 %
CEP13270	94 %	34 %
CEP18248	0 %	0 %

Tabelle 3.15: Prozent an enzymatisch aktiven Heterodimeren aus TSP1- und TSP2-Proteinen aus *Parasitella parasitica*. Aktivität wird über die Zugänglichkeit des aktiven Zentrums für Cosubstrat und Substrat definiert. Zeilen zeigen den Rezeptor, Spalten den Liganden des Protein-Protein-Dockings.

Abbildung 3.46 zeigt einige der modellierten Dimere aus TSP1- und TSP2-Proteinen aus *Parasitella parasitica*. Die aktiven Zentren der TSP1-Proteine blieben in den meisten Fällen offen zugänglich. Die Inaktivierung einzelner TSP1-Proteine kam durch eine Stauung der Proteinstruktur und damit verbundenen Verengung der Bindetaschen zustande.

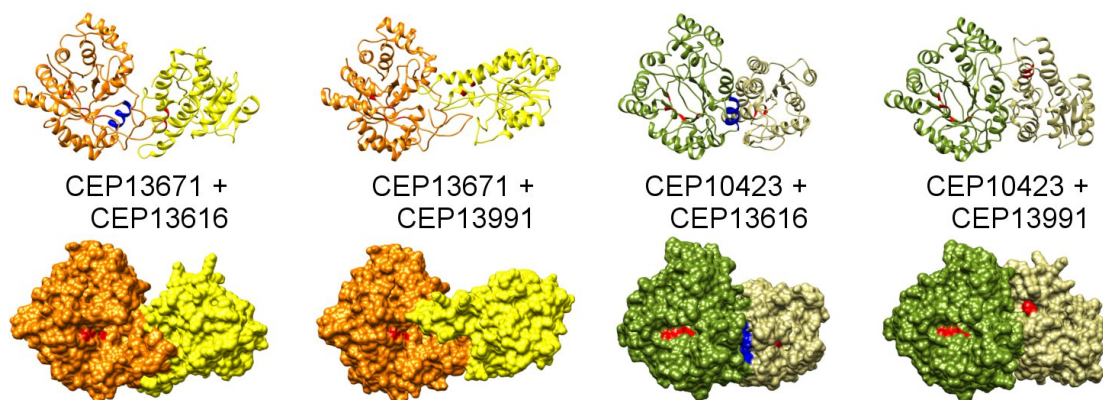


Abbildung 3.46: Modellierung der TSP1-TSP2-Dimere aus *Parasitella parasitica* aus den TSP1-Proteinen CEP13671 und CEP10423 als Rezeptor und den TSP2-Proteinen CEP13616 und CEP13991 als Ligand. Die aktiven Zentren sind rot, die α 5-Helices blau markiert.

Tabelle 3.16 zeigt die Ergebnisse aus dem zweiten Protein-Protein-Docking mit TSP2 als Rezeptor und TSP1 als Ligand. Die beiden TSP2-Varianten unterschieden sich deutlich in ihrer enzymatischen Aktivität. CEP13991 war nur in Kombination mit CEP14056 enzymatisch aktiv. CEP13671 hemmte die enzymatische Aktivität von CEP13616.

	CEP13671	CEP14056	CEP10423	CEP10425	CEP13270	CEP18248
CEP13616	78 %	100 %	100 %	100 %	95 %	88 %
CEP13991	0 %	52 %	0 %	0 %	0 %	0 %

Tabelle 3.16: Prozent an enzymatisch aktiven Heterodimeren aus TSP2- und TSP1-Proteinen aus *Parasitella parasitica*. Aktivität wird über die Zugänglichkeit des aktiven Zentrums für Cosubstrat und Substrat definiert. Zeilen zeigen den Rezeptor, Spalten den Liganden des Protein-Protein-Dockings.

Abbildung 3.47 zeigt einige der modellierten Dimere aus TSP2- und TSP1-Proteinen aus *Parasitella parasitica*. Die aktiven Zentren der TSP2-Proteine wurden bei der Inaktivierung nur in Kombination mit dem Original-TSP1-Protein CEP13616 von der flexiblen Schleife des Partnerproteins verschlossen. In den meisten Fällen kam die Inaktivierung durch eine Stauchung der Proteinstruktur zustande.

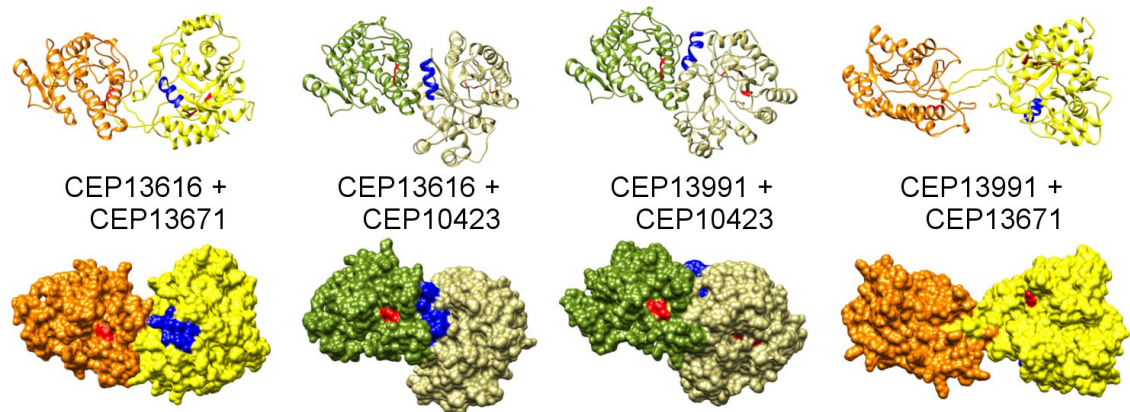


Abbildung 3.47: Modellierung der TSP2-TSP1-Dimere aus den TSP2-Proteinen CEP13616 und CEP13991 als Rezeptor und den TSP1-Proteinen CEP13671 und CEP10423 als Ligand. Das aktive Zentrum der Enzyme ist rot, die α 5-Helices blau markiert.

3.4.5 Modellierung von TSP1-Dimeren zwischen *Absidia glauca* und *Parasitella parasitica*

An der Schnittstelle zwischen Wirt und Parasit kommt es zur Plasmafusion. Dabei kommen TSP1-Proteine aus *Absidia glauca* und *Parasitella parasitica* miteinander in Kontakt. Um herauszufinden, wie die enzymatische Aktivität der TSP1-Proteine eines Pilzes von den TSP1-Proteinen des Interaktionspartners beeinflusst wird, wurden Dimere mit TSP1-Proteinen aus beiden Zygomyceten modelliert und auf ihre Aktivität überprüft.

Da die Dimere nicht symmetrisch sind, wurden alle Proteine in den Kombinationen zweimal gestartet, einmal als Rezeptor und einmal als Ligand. Tabelle 3.17 zeigt die Ergebnisse aus dem Protein-Protein-Docking mit den vier TSP1-Proteinen aus *Absidia glauca* als Rezeptor und den sechs TSP1-Proteinen aus *Parasitella parasitica* als Ligand. TSP1-Proteine aus *Absidia glauca* wurden von der flexiblen Schleifenregion des TSP1-Proteins CEP13671 aus *Parasitella parasitica* inaktiviert. Besonders die Kombination aus den beiden Original-TSP1-Proteinen SAM02260 aus *Absidia glauca* und CEP13671 aus *Parasitella parasitica* ergab inaktive Heterodimere. Das Trisporoid fand keinen Zugang zum aktiven Zentrum des Enzyms. Das Enzym war inaktiv. Dimere, die mit den TSP1-Proteinen CEP14056, CEP10423 und CEP10425 aus *Parasitella parasitica* gebildet wurden, waren immer enzymatisch aktiv. Das TSP1-Protein CEP13270 aus *Parasitella parasitica* war mit den beiden *Absidia*-spezifischen TSP1-Proteinen besonders aktiv. Die TSP1-Proteine des ursprünglichen Typs wurden von diesem TSP1-Protein inhibiert.

	CEP13671	CEP14056	CEP10423	CEP10425	CEP13270	CEP18248
SAL98187	45 %	100 %	100 %	100 %	100 %	100 %
SAL98947	6 %	100 %	100 %	100 %	66 %	80 %
SAM02260	5 %	100 %	100 %	100 %	38 %	69 %
SAM06439	46 %	100 %	100 %	100 %	100 %	87 %

Tabelle 3.17: Prozent an enzymatisch aktiven Dimeren aus TSP1-Proteinen aus *Absidia glauca* und *Parasitella parasitica*. Aktivität wird über die Zugänglichkeit des aktiven Zentrums für Cosubstrat und Substrat definiert. Zeilen zeigen den Rezeptor, Spalten den Liganden des Protein-Protein-Dockings.

Abbildung 3.48 zeigt einige der modellierten Dimere aus TSP1-Proteinen aus *Absidia glauca* und *Parasitella parasitica*. Die aktiven Zentren der TSP1-Proteine wurden bei der Inaktivierung von der flexiblen Schleife des Partnerproteins verschlossen. Bei aktiven Dimeren ist die $\alpha 5$ -Helix an der Bildung der Dimerisierungsschnittstelle beteiligt.

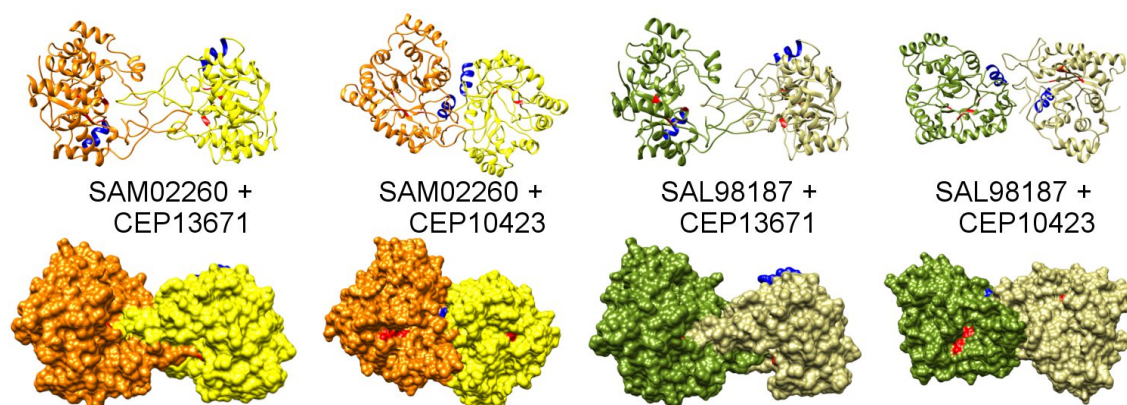


Abbildung 3.48: Modellierung der TSP1-Dimere aus einem Original-TSP1 SAM02260, dem *Absidia*-spezifischen TSP1 SAM98187, dem Original-TSP1 CEP13671 und dem *Parasitella*-spezifischen TSP1 CEP10423. Die aktiven Zentren sind rot, die $\alpha 5$ -Helices blau markiert. Die aktiven Zentren des Heterodimers aus den beiden Original-TSP1-Proteinen SAM02260 und CEP13671 werden von den flexiblen Schleifenstrukturen der Partner blockiert. Das Dimer wird auf diese Weise inaktiviert.

Tabelle 3.18 zeigt die Ergebnisse aus dem Protein-Protein-Docking mit den sechs TSP1-Proteinen aus *Parasitella parasitica* als Rezeptor und den vier TSP1-Proteinen aus *Absidia glauca* als Ligand. Die TSP1-Proteine CEP14056, CEP13270 und CEP18248 aus *Parasitella parasitica* sind regulatorische Untereinheiten. Sie zeigten keine Enzymaktivität, unabhängig von dem gewählten Partnerprotein aus *Absidia glauca*, sondern aktivierten andere TSP1-Proteine. Das TSP1-Protein CEP13671 wurde von den TSP1-Proteinen aus *Absidia glauca* inhibiert. Das einzige TSP1-Protein, das in Kombination mit allen TSP1-Proteinen aus *Absidia glauca* zu 100 % enzymatisch aktiv war, ist CEP10423. Das TSP1-Protein CEP10425 zeigte ebenfalls eine hohe enzymatische Aktivität.

Abbildung 3.49 zeigt einige der modellierten Dimere aus TSP1-Proteinen aus *Parasi-*

Aktivierung von TSP1-Homodimeren

	SAL98187	SAL98947	SAM02260	SAM06439
CEP13671	20 %	21 %	19 %	27 %
CEP14056	0 %	0 %	0 %	0 %
CEP10423	100 %	100 %	100 %	100 %
CEP10425	87 %	93 %	100 %	91 %
CEP13270	0 %	0 %	0 %	0 %
CEP18248	0 %	0 %	0 %	0 %

Tabelle 3.18: Prozent an enzymatisch aktiven Dimeren aus TSP1-Proteinen aus *Parasitella parasitica* und *Absidia glauca*. Aktivität wird über die Zugänglichkeit des aktiven Zentrums für Cosubstrat und Substrat definiert. Zeilen zeigen den Rezeptor, Spalten den Liganden des Protein-Protein-Dockings.

tella parasitica und *Absidia glauca*. Die aktiven Zentren der TSP1-Proteine wurden bei der Inaktivierung von der flexiblen Schleife des Partnerproteins verschlossen. Bei aktiven Dimeren ist die $\alpha 5$ -Helix an der Bildung der Dimerisierungsschnittstelle beteiligt. Das Heterodimer aus CEP10423 und SAM02260 ähnelt dem Dimer der Xylose-Reduktase, die als Vorlagenstruktur verwendet wurde.

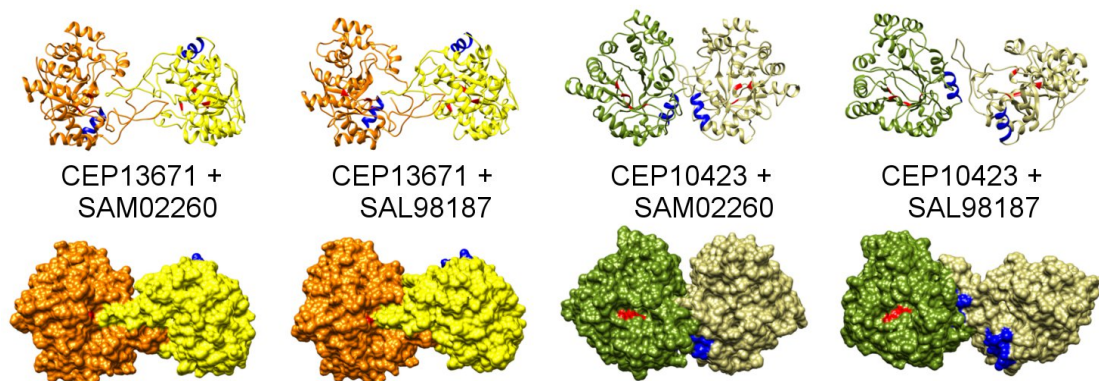


Abbildung 3.49: Modellierung der TSP1-Dimere aus dem Original-TSP1 CEP13671, dem *Parasitella*-spezifischen TSP1 CEP10423, dem Original-TSP1 SAM02260 und dem *Absidia*-spezifischen TSP1 SAM98187. Die aktiven Zentren sind rot, die $\alpha 5$ -Helices blau markiert.

3.5 *In silico* Mutanten von TSP1 aus *Parasitella parasitica*:

Aktivierung von TSP1-Homodimeren

Um die Stabilität der aktiven und inaktiven Homodimere zu untersuchen, wurde eine Reihe von *in silico* Mutanten der Original-TSP1-Sequenz CEP13671 aus *Parasitella parasitica* erzeugt. Es wurden entscheidende Abweichungen in der Proteinstruktur ermittelt, die zu einer Inaktivierung des Partnerproteins bei der Bildung von Homodimeren aus CEP13761 führten.

Der Vergleich der beiden TSP1-Strukturen CEP13671 und CEP10423 zeigt, dass sich die entscheidenden Unterschiede auf drei Bereiche verteilen: (i) eine innere Schleifenre-

Aktivierung von TSP1-Homodimeren

gion, (ii) die $\alpha 5$ -Helix und (iii) die C-terminale Schleifenregion. In Abbildung 3.50 sind diese drei Bereiche in einem Strukturalignment der beiden TSP1-Varianten dargestellt.

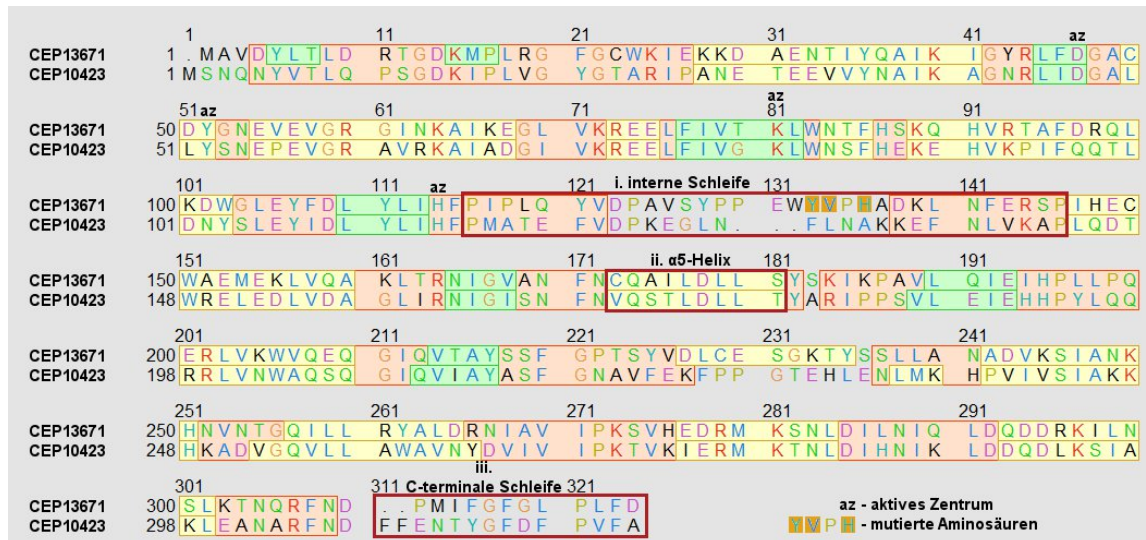


Abbildung 3.50: Strukturalignment der beiden entscheidenden TSP1-Varianten aus *Parasitella parasitica*. Die Proteinstrukturen des aktivierenden (CEP10423) und des inaktivierenden (CEP13671) TSP1-Monomers unterscheiden sich vor allem in der zentralen Schleifenregion. Diese ist bei dem inaktivierenden TSP1-Protein drei Aminosäuren länger, als bei der anderen Variante. Ein Austausch der drei orange markierten Aminosäuren genügt, um das inaktivierende in ein aktivierendes TSP1-Monomer umzuwandeln.

Abbildung 3.51 zeigt den Einfluss der oben beschriebenen Strukturelemente auf das Dimerisierungsverhalten. Der Austausch einzelner Regionen führte bei der inneren Schleifenregion zu einer Änderung der Enzymaktivität der aus den Mutanten bestehenden Homodimere hin zu aktiven Enzymkomplexen. Der einzelne Austausch der $\alpha 5$ -Helix und der C-terminalen Schleife ergaben inaktive Homodimere, die dem Ausgangsprotein ähnelten. Der gleichzeitige Austausch aller drei Elemente führte zu aktiven Dimeren, die dem aktiven Homodimer aus CEP10423 ähnelten.

Erstaunlicherweise genügte bereits der Austausch von drei Aminosäuren (Y132E, V133E, H135E) innerhalb der inneren Schleifenregion, um aktive Homodimere entstehen zu lassen. Zum Vergleich ist das aktive Homodimer aus dem TSP1-Protein CEP41023 abgebildet.

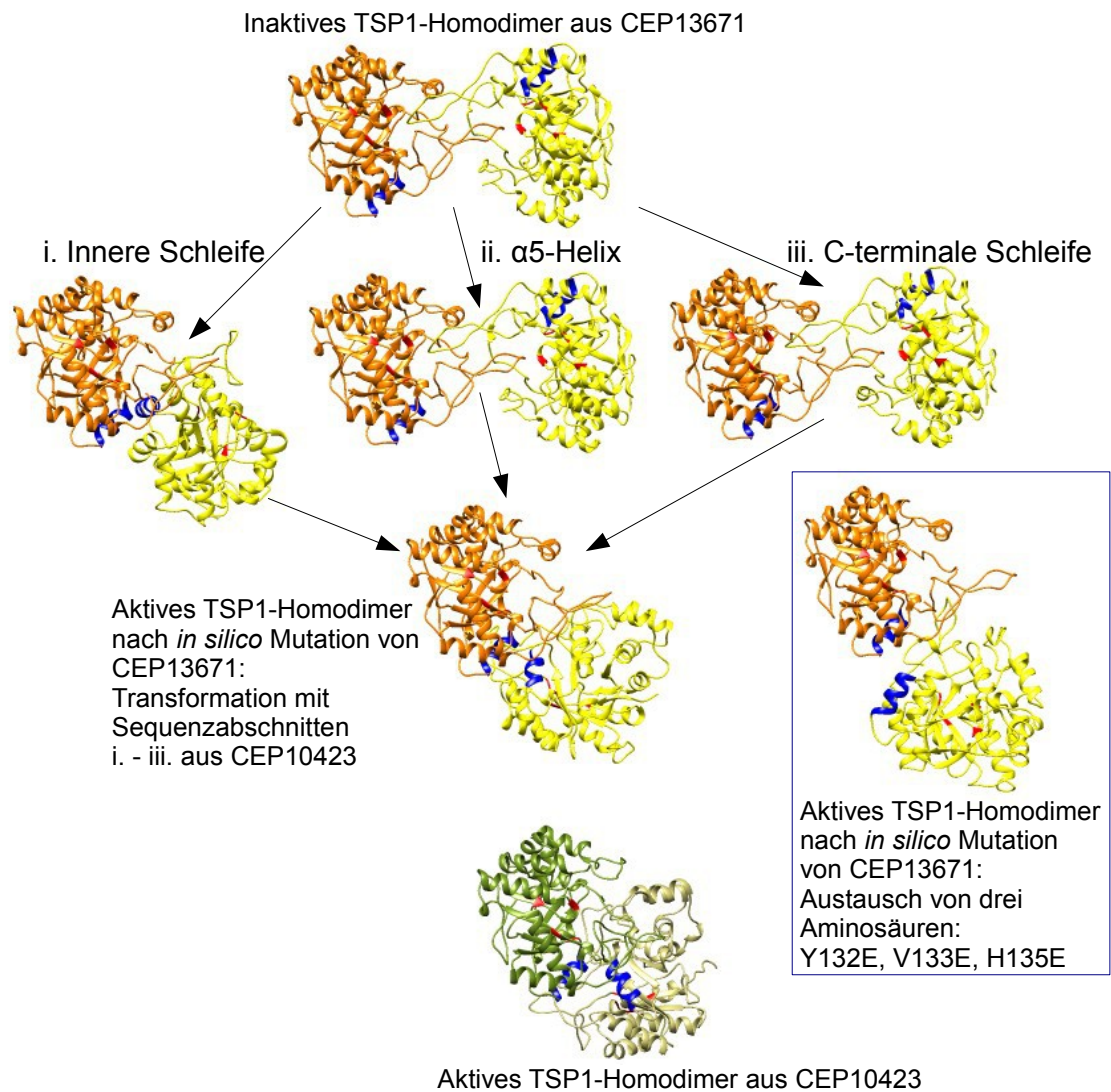


Abbildung 3.51: Modellierung der TSP1-Proteine aus *Parasitella parasitica* nach Veränderung (Mutation) der Proteinsequenz.

Die Schleifenregionen sind sehr variabel zwischen den verschiedenen TSP1-Isoformen. TSP1-Mutanten wurden mit dem Austausch kompletter Regionen und einzelner Aminosäuren erzeugt. Erstaunlicherweise genügt der Austausch von nur drei Aminosäuren für eine Konformationsänderung, die eine Aktivierung des Homodimers bewirkt.

3.6 Detektion von horizontalem Gentransfer zwischen Bakterien und Protozoen

Horizontaler Gentransfer (HGT) zwischen verschiedenen Bakterien wurde bereits sehr intensiv untersucht. Über den Gentransfer zwischen Bakterien und Eukaryoten ist weniger bekannt. Um jüngere Gentransferereignisse, zum Beispiel von Bakterien zu Protozoen, zu finden, braucht es spezifischere Programme, die mit kleinen Datensätzen arbeiten können. Für genau dieses Problem wurde ein Algorithmus entwickelt, der vier verschiedene Ansätze zur Detektion von horizontalem Gentransfer in einer bewertungsbasierten Java-Anwendung kombiniert. Der Algorithmus wurde in *Endocytobiosis and Cell Research* (Ellenberger, Siegmund et al., 2016) publiziert. Der HGT Calculator wurde an Genen aus Trypanosomatiden getestet.

3.6.1 Gezielte Identifizierung von HGT in Isocitrat-Dehydrogenasen

Der Arbeitsablauf des HGT Calculator (siehe Abschnitt 3.6) wird im Folgenden am Beispiel der Isocitrat-Dehydrogenase erläutert. Die Gene der Isocitrat-Dehydrogenase wurden als Anschauungsbeispiel gewählt, weil es sich hierbei um einen essentiellen Teil des Citratzyklus in Prokaryoten und Eukaryoten handelt und die Gene daher für alle Organismen des Testdatensatzes verfügbar sind.

Der Datensatz wurde mit Hilfe des HGT Downloaders, einem Teilprogramm des HGT-Calculators, zusammengestellt und bearbeitet. Geeignete Sequenzen wurden über eine KEGG-Orthologie-Identifikationsnummer (www.genome.jp/kegg, K00031) gefunden. Für *Strigomonas culicis* und *Trypanosoma rangeli* gab es keine Einträge in KEGG. Sequenzen aus diesen beiden Organismen wurden mit einer Textsuche in UniProt gefunden. 12 eukaryotische und 16 prokaryotische Sequenzen wurden ausgewählt und aufgrund ihrer Proteindomänen in drei Gruppen sortiert. Bereits bei der Auswahl der Sequenzen fiel auf, dass *Leishmania major* und *Strigomonas culicis* zwei verschiedene Isocitrat-Dehydrogenase-Gene besitzen. *Trypanosoma rangeli* besitzt nur ein solches Gen. Nach der Eingruppierung der ausgewählten Sequenzen wurden diese beiden Varianten als ICDH_1 (UniProt: Q4Q3T2, S9UPN9) und ICDH_2 (UniProt: Q4QHI7, S9UPW7) gekennzeichnet. Die Sequenz aus *Trypanosoma rangeli* gehört in die zweite Gruppe (ICDH_2, UniProt: A0A061IWC2). Die beiden Genvarianten besitzen unterschiedliche Proteindomänen. Tabelle 3.19 zeigt, wie unterschiedlich diese Gene sind, obwohl sie zu derselben KEGG-Orthologie (K00031) gehören.

In ICDH_1 wies eine der Proteindomänen bereits auf ein HGT-Ereignis hin. Die Domäne IPR004439 (Isocitrate_DH_NADP_dimer_prok) bezeichnet eine bakterielle NADP-abhängige Isocitrat-Dehydrogenase, die Dimere bildet. Der Anteil an Spezies, bei denen diese Proteindomäne vorkommt, zeigte deutlich, dass man diese nicht in einem Eukaryoten erwarten würde. 3 510 Bakterien besaßen diese Proteindomäne. Bei Eukaryoten war

	Block A Proteindomänen mit InterPro-Identifikationsnummern	Block B Anteil prokaryotischer BLAST_Treffer Alien-Index ≥ 45	Block C Ungewöhnlicher GC-Gehalt und niedriger Codon Adaptation Index
<i>Leishmania major</i>			
ICDH_1 UniProt ID: Q4Q3T2	IPR019818 IsoCit/isopropylmalate_DH_CS. IPR001804 Isocitrate/isopropylmalate_DH. IPR004439 Isocitrate_DH_NADP_dimer_prok. IPR024084 IsoPropMal-DH-like_dom.	alle: 250 Archaeen: 0 Bakterien: 204 Eukaryoten: 46 <i>L. major</i> : 1 AI: 200 (Proteobakterien)	GC-Gehalt kodierender Bereich: 59.94 % erste Codon-Position: 59.15 % zweite Codon-Position: 42.96 % dritte Codon-Position: 77.7 % CAI: 0.705
ICDH_2 UniProt ID: Q4QHI7	IPR019818 IsoCit/isopropylmalate_DH_CS. IPR004790 Isocitrate_DH_NADP. IPR024084 IsoPropMal-DH-like_dom.	alle: 250 Archaeen: 0 Bakterien: 31 Eukaryoten: 219 <i>L. major</i> : 1 AI: 0	GC-Gehalt kodierender Bereich: 59.48 % erste Codon-Position: 54.59 % zweite Codon-Position: 38.07 % dritte Codon-Position: 85.78 % CAI 0.783
<i>Strigomonas culicis</i>			
ICDH_1 UniProt ID: S9UPN9	IPR019818 IsoCit/isopropylmalate_DH_CS. IPR001804 Isocitrate/isopropylmalate_DH. IPR004439 Isocitrate_DH_NADP_dimer_prok. IPR024084 IsoPropMal-DH-like_dom.	alle: 250 Archaeen: 0 Bakterien: 240 Eukaryoten: 10 <i>S. culicis</i> : 1 AI: 200 (Proteobakterien)	GC-Gehalt kodierender Bereich: 67.14 % erste Codon-Position: 64.22 % zweite Codon-Position: 44.55 % dritte Codon-Position: 92.65 % CAI 0.859
ICDH_2 UniProt ID: S9UPW7	IPR019818 IsoCit/isopropylmalate_DH_CS. IPR004790 Isocitrate_DH_NADP. IPR024084 IsoPropMal-DH-like_dom.	alle: 250 Archaeen: 0 Bakterien: 22 Eukaryoten: 228 <i>S. culicis</i> : 2 AI: 0	GC-Gehalt kodierender Bereich: 63.3 % erste Codon-Position: 58.62 % zweite Codon-Position: 38.39 % dritte Codon-Position: 92.87 % CAI 0.885
<i>Trypanosoma rangeli</i>			
ICDH_2 UniProt ID: A0A061IWC2	IPR019818 IsoCit/isopropylmalate_DH_CS. IPR004790 Isocitrate_DH_NADP. IPR024084 IsoPropMal-DH-like_dom.	alle: 250 Archaeen: 0 Bakterien: 26 Eukaryoten: 224 <i>T. rangeli</i> : 1 AI: 0	GC-Gehalt kodierender Bereich: 56.57 erste Codon-Position: 57.11 zweite Codon-Position: 40.37 dritte Codon-Position: 72.25 CAI 0.818
HGT in ICDH_1 wahrscheinlich?	ja	ja	nein

Tabelle 3.19: Die parametrische Analyse belegt den horizontalen Gentransfer der ICDH_1-Variante in *Leishmania major* und *Strigomonas culicis*. Kriterien, die für ein HGT-Ereignis sprechen, sind fett gedruckt, zum Beispiel prokaryotische Proteindomänen, ein hoher Alien-Index und ein abweichender GC-Gehalt. In Klammern steht der mögliche Ursprung der transferierten Gene. (Tabelle nach Ellenberger, Siegmund et al. (2016))

sie nur 16 mal zu finden.

Vier dieser Proteine stammten von *Leishmania*-Spezies und eine aus *Strigomonas culicis*.

Die Listen der BLAST-Ergebnisse unterschieden sich deutlich zwischen den beiden Isocitrat-Dehydrogenasen aus *Leishmania major* und *Strigomonas culicis*. Sie zeigten ein HGT-Ereignis deutlich an. Die Sequenz der ICDH_1-Variante aus *Leishmania major* lieferte 204 bakterielle und nur 46 eukaryotische BLAST-Treffer. Diese eukaryotischen Sequenzen stammten aus verschiedenen *Leishmania*-Spezies, *Angomonas deanei*, und *Strigomonas culicis*. Die zweite Sequenz aus *Leishmania major*, ICDH_2, lieferte hingegen nur 31 bakterielle Sequenzen. 20 der insgesamt 219 eukaryotischen BLAST-Treffer stammten aus *Leishmania*-Spezies. Die Liste beinhaltete auch Sequenzen aus *Angomonas deanei*, *Strigomonas culicis*, *Trypanosoma*, *Paramecium*, *Toxoplasma*, *Saccharomyces*, *Arabidopsis*, und *Homo*. In der Liste waren keine weiteren Sequenzen aus *Leishmania major* zu finden, außer der Ausgangssequenz selbst. Die Ergebnisse der BLAST-Suchen für die beiden Sequenzen aus *Strigomonas culicis* waren vergleichbar. Zehn der insgesamt 250 BLAST-Treffer für die ICDH_1-Sequenz stammten von Eukaryoten, während es für die ICDH_2-Sequenz 228 eukaryotische BLAST-Treffer gab. ICDH_2 aus *Trypanosoma rangeli* lieferte ebenfalls deutlich mehr eukaryotische BLAST-Treffer. Die BLAST-Suche resultierte in 224 eukaryotischen und nur 26 bakteriellen BLAST-Treffern.

Der angepasste Alien-Index (AI) zeigte den Unterschied zwischen den beiden Varianten der Isocitrat-Dehydrogenase deutlich. ICDH_1-Sequenzen aus *Leishmania major* und *Strigomonas culicis* ergaben einen Alien-Index von 200. Das ist der höchste mögliche Wert für einen Alien-Index. Die Werte für die meisten ICDH_2-Sequenzen waren hingegen null. Der Alien-Index für ICDH_2 aus *Paramecium tetraurelia* war sogar -200, der kleinste mögliche Alien-Index. Der AI-Ansatz belegte den horizontalen Gentransfer von ICDH_1 eindeutig. Aus den BLAST-Ergebnissen lies sich ein Proteobakterium als möglicher Ursprung des ICDH_1-Gens ableiten. Sie scheinen das Ergebnis eines multiplen HGT-Ereignisses zwischen einem Vorfahren von *Gilliamella* und Trypanosomatiden zu sein. Der beste prokaryotische BLAST-Treffer für diese Gene aus *Leishmania major* und *Strigomonas culicis* stammte aus einem *Gilliamella*-Stamm mit Sequenzidentitäten von 67 % und 68 %.

Die Ergebnisse der Analyse auf Nukleotidebene widersprachen den bereits gewonnenen Hinweisen auf ein HGT-Ereignis in ICDH_1. Abweichungen im GC-Gehalt in *Leishmania major* deuteten eher auf ICDH_2 hin. Für diese Sequenz wich der GC-Gehalt für alle drei Codon-Positionen vom Spezies-spezifischen Durchschnitt ab. Für ICDH_1, wies nur der GC-Gehalt der zweiten Codon-Position auf eine Abweichung hin. Dieses unerwartete Ergebnis bestätigte Beobachtungen, wonach diese Methoden zum Auffinden von HGT-Ereignissen inkonsistent mit anderen Anwendungen ist (Ragan, 2001). Ein niedriger Codon Adaptation Index weist auf ein HGT-Ereignis hin. In einigen Protozoen

sind die Werte für den Codon Adaptation Index allerdings generell gering, unabhängig davon, ob es sich um ein HGT-Ereignis handelt oder nicht. Diese niedrigen Werte sind ein Ausdruck niedriger Gen-Expressionsraten. Ursache hierfür ist möglicherweise ein parasitischer Lebensstil. Der Unterschied beim Codon Adaptation Index der beiden Isocitrat-Dehydrogenase-Sequenzen aus *Leishmania major* war zu gering, um aussagekräftig zu sein. Trotzdem war der Codon Adaptation Index für ICDH_1 (0.705) geringer als für ICDH_2 (0.783). Im Fall von *Strigomonas culicis* und *Trypanosoma rangeli* war der Codon Adaptation Index nur eingeschränkt nutzbar, da die Codon-Tabellen aus der Datenbank nicht aussagekräftig genug waren. Tabelle 3.19 zeigt die Ergebnisse der parametrischen Analysen der Isocitrat-Dehydrogenase-Sequenzen aus *Leishmania major*, *Strigomonas culicis* und *Trypanosoma rangeli*.

Die Multi-Level-Analyse der Neighbor-Joining-Bäume belegte, dass es sich bei der ICDH_1-Variante um ein HGT-Ereignis handelt. Die Sequenz für ICDH_1 aus *Leishmania major* lag im Level1-Baum direkt neben einer bakteriellen Sequenz und das mit einer Distanz, die größer war als der vorgegebene Schwellwert. Im Level2-Baum bildeten die beiden ICDH_1-Sequenzen von *Leishmania major* und *Strigomonas culicis* zusammen mit der bakteriellen Sequenz eine Gruppe und das ebenfalls mit einer Distanz, die größer war als der vorgegebene Schwellwert. Auch im Level3-Baum zeigte sich deutlich, dass es sich bei ICDH_1 um eine HGT-Ereignis handelt. Der Level3-Baum lies sich in einen eukaryotischen und einen prokaryotischen Teilbaum zerlegen. Das beweist die Korrektheit der zugrundeliegenden Baumtopologie. Die ICDH_1-Sequenzen von *Leishmania major* und *Strigomonas culicis* waren die einzigen eukaryotischen Sequenzen innerhalb des prokaryotischen Teilbaums. Die ICDH_2-Sequenzen von *Leishmania major*, *Strigomonas culicis* und *Trypanosoma rangeli* bildeten zusammen mit den übrigen eukaryotischen Sequenzen den eukaryotischen Teilbaum. Abbildung 2.14 zeigt die Bäume aller drei Level für die Analyse der Isocitrat-Dehydrogenase in Trypanosomatiden.

Abbildung 3.52 zeigt die Berechnung der HGT-Werte für die Isocitrat-Dehydrogenase-Gene aus *Leishmania major*. Der HGT-Wert für die ICDH_1-Variante aus *Leishmania major* war 55, der für die ICDH_1-Variante aus *Strigomonas culicis* 70. Alle anderen HGT-Werte, inklusive der für die ICDH_2-Variante aus *Trypanosoma rangeli*, waren null. Entscheidend sind dafür der geringe Anteil prokaryotischer Sequenzen innerhalb der BLAST-Ergebnisse und die Lage in den Neighbor-Joining-Bäumen.

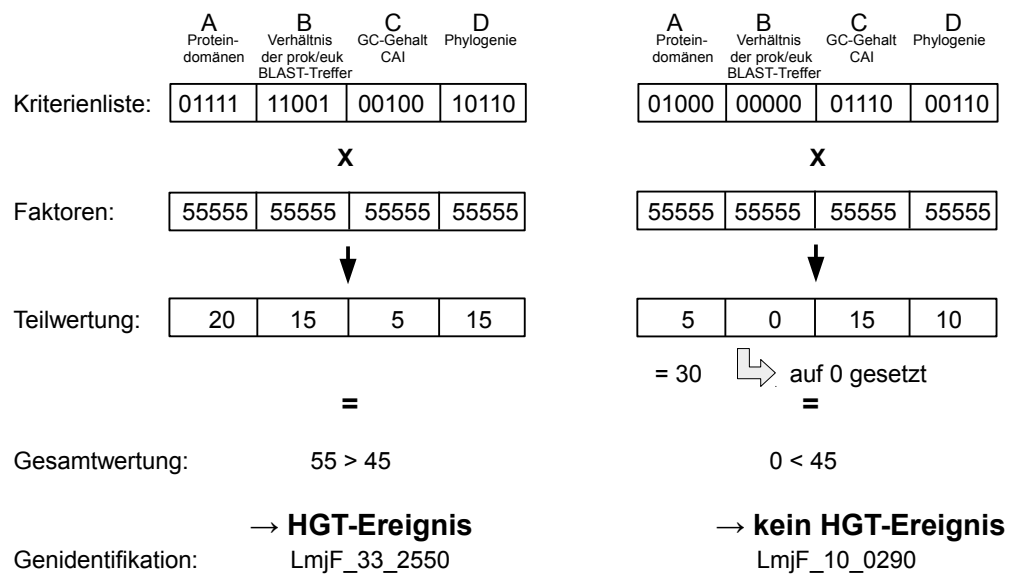


Abbildung 3.52: HGT-Werte für Isocitrat-Dehydrogenasen aus *Leishmania major*.

Der HGT-Wert eines Gens setzt sich aus den Ergebnissen der einzelnen Analysen zusammen. Den einzelnen Kriterien lassen sich unterschiedliche Gewichte zuweisen. Die HGT-Werte für Standardgewichtung der beiden Gene unterschieden sich deutlich und zeigten, dass es sich bei einer der beiden Isocitrat-Dehydrogenasen um ein HGT-Ereignis handelt (LmjF_33_2550). (Tabelle nach Ellenberger, Siegmund et al. (2016))

3.6.2 Gezielte Identifizierung von HGT in *Trypanosoma rangeli*

Gene aus *Trypanosoma rangeli* zeigten ähnliche Signale für HGT-Ereignisse, wie solche aus *Trypanosoma brucei* und *Trypanosoma cruzi*. Alle HGT-Werte, die mit HGT Calculator berechnet wurden, sind in Tabelle 3.20 zusammengefasst.

Gene für Argininosuccinat-Synthase, Ferrochelatase und Homoserin-Dehydrogenase gehen in Trypanosomatiden auf ein altes HGT-Ereignis zurück (Alves et al., 2013; Operdoes und Michels, 2007). Die Gene wurden von einem Bakterium in den gemeinsamen Vorfahren von Trypanosomatiden mit und ohne Endosymbionten überführt. *Trypanosoma*-Spezies scheinen diese Gene verloren zu haben. Daher gibt es keine solchen Gene in *Trypanosoma rangeli*.

Die Gene der Fumarat-Hydratase werden als multiples HGT-Ereignis beschrieben (Operdoes und Michels, 2007). Das Gen kommt sowohl in Trypanosomatiden mit und ohne Endosymbionten, als auch in einigen Apicomplexa vor. Der HGT-Wert für FUM_2 aus *Trypanosoma rangeli* (HGT-Wert: 35) war zu niedrig, um als rezentes HGT-Ereignis gewertet zu werden. Der Alien-Index war null. Allerdings bestätigte die Sequenzanalyse den horizontalen Gentransfer.

	ASS_1	FECH_1	FUM_1	FUM_2	GDH_1	HDH_1	HDH_2	ICDH_1	ICDH_2	IPPI_1	IPPI_2	METE_1	SAT_1	XYLB_1
<i>Tetrahymena</i>		0 b t	0 b t		0 b t				0 b t		0 b t			
<i>Paramecium</i>		0 b t	0 b t		0 b t				0 b t		0 b t			
<i>Strigomonas</i>	0 (40) t → Firmicutes	0 (40) t → Proteo- bakterien		55	0 (35) t → Proteo- bakterien	65 → Proteo- bakterien	0 (40) t → Firmicutes	70 → Proteo- bakterien	0 b t	65 → Cyano- bakterien		0 (45) t → Proteo- bakterien	40 → Bacteroi- detes	
<i>Leishmania</i>	0 (40) t → Firmicutes	0 (25) t → Proteo- bakterien		45	0 (30) t → Proteo- bakterien		0 (30) t → Firmicutes	55 → Proteo- bakterien	0 b t	45 → Cyano- bakterien		0 (30) t → Proteo- bakterien	35 → Bacteroi- detes	55
<i>Trypanosoma</i>				35	0 (25) t → Proteo- bakterien				0 b t	35 → Cyano- bakterien		0 (30) t → Proteo- bakterien	25 → Bacteroi- detes	40
<i>Toxoplasma</i>		0 (20) t		25	0 (5) t				0 b t					
<i>Plasmodium</i>		0 (25) t		20	0 (25) t				0 b t					
<i>Saccharomyces</i>	0 b t	0 b t	0 b t		0 b t		0 b t		0 b t		0 b t	0 b t		0 b t
<i>Arabidopsis</i>	0 (25) t	0 b t	0 b t		0 (20) t → Proteo- bakterien		0 (20) t		0 b t		0 b t	0 (20) t → Bakterien	0 b t	0 b t
<i>Homo</i>	0 b t	0 b t	0 b t		0 b t				0 b t		0 b t			0 b t

Tabelle 3.20: HGT-Werte für HGT-Ereignisse in Trypanosomatiden.

Darstellung in der Tabelle: HGT-Wert (HGT-Zwischenwert)

b: Die BLAST-Ergebnisse sind nicht überzeugend.

t: Die Topologie der Neighbor-Joining-Bäume liefert keine Belege für horizontalen Gentransfer.

(Tabelle nach Ellenberger, Siegmund et al. (2016))

Glutamat-Dehydrogenasen werden in unterschiedliche Klassen unterteilt. Das HGT-Gen aus *Leishmania major* gehört zu einer anderen Klasse, als das Glutamat-Dehydrogenase-Gen aus *Trypanosoma rangeli*. Deshalb wurde stattdessen ein passenderes Gen aus *Trypanosoma cruzi* ausgewählt, um einen möglichen horizontalen Gentransfer zu identifizieren. Diese Klasse der Glutamat-Dehydrogenasen wird in der Literatur als multiples HGT-Ereignis beschrieben (Alves et al., 2013; Oppendoes und Michels, 2007). Der HGT-Wert war zu niedrig, um als rezentes HGT-Ereignis gewertet zu werden (HGT-Wert: 25). Der Alien-Index war 200. Allerdings lieferte die Sequenzanalyse aufgrund von weiteren HGT-Ereignissen in multi-zellulären Eukaryoten kein positives HGT-Signal.

Isopentenylidiphosphat-Delta-Isomerase kommt in Trypanosomatiden mit und ohne eigene Endosymbionten vor. Dieses Gen ist ein HGT-Ereignis (Oppendoes und Michels, 2007). Das Gen stammt ursprünglich aus einem Cyanobakterium. Die IPPI_1-Variante beinhaltete eine prokaryotische Proteindomäne und lieferte eine große Anzahl prokaryotischer BLAST-Treffer. Der HGT-Wert identifizierte dieses multiple HGT-Ereignis in *Trypanosoma rangeli* (HGT-Wert: 35).

5-Methyltetrahydropteroyltriglutamat-Homocystein-Methyltransferase kommt sowohl in Trypanosomatiden mit und ohne Endosymbionten vor. Allerdings gibt es keine Sequenz für *Trypanosoma rangeli*. Deshalb wurde stattdessen ein passendes Gen aus *Trypanosoma brucei* ausgewählt, um einen möglichen horizontalen Gentransfer zu identifizieren. Die HGT-Werte waren zu niedrig, um als rezentes HGT-Ereignis gewertet zu werden (HGT-Wert: 30). Trotzdem identifizierte der HGT-Wert dieses multiple HGT-Ereignis (Alves et al., 2013; Oppendoes und Michels, 2007) in *Trypanosoma brucei*.

Serin-O-Acetyltransferasen gehen in Trypanosomatiden auf ein altes HGT-Ereignis zurück (Alves et al., 2013; Oppendoes und Michels, 2007). Das Gen wurde von einem

Bakterium aus der Gruppe der Bacteroidetes in den gemeinsamen Vorfahren von Trypanosomatiden mit und ohne Endosymbionten überführt. Allerdings gibt es keine Sequenz für *Trypanosoma rangeli*. Deshalb wurde stattdessen ein passendes Gen aus *Trypanosoma cruzi* ausgewählt, um einen möglichen horizontalen Gentransfer zu identifizieren. Die HGT-Werte waren zu niedrig, um als rezentes HGT-Ereignis gewertet zu werden (HGT-Wert: 25). Trotzdem identifizierte der HGT-Wert dieses multiple HGT-Ereignis in *Trypanosoma cruzi*.

Das Gen für Xylulokinase ist ein HGT-Ereignis in *Leishmania major* (Opperdoes und Michels, 2007) und *Trypanosoma rangeli*. Der HGT-Wert für *Trypanosoma rangeli* war 40 und identifizierte dieses multiple HGT-Ereignis. Der Grund für den Alien-Index von null liegt in einer Sequenz aus *Oryza sativa* (UniProt: B8BME9), die scheinbar auf eine Sequenz aus *Enterobacter* (UniProt: V3RGC1) zurückgeht oder nicht Teil des Genoms ist. Ohne diesen BLAST-Treffer wäre der Alien-Index der Xylulokinase aus *Trypanosoma rangeli* 200 und würde das HGT-Ereignis belegen.

4 Diskussion

4.1 Vergleich der Genome von Wirt und Parasit

4.1.1 Kerngenom

Das Genom von *Absidia glauca* (49 Mbp) ist etwas größer als das Genom von *Parasitella parasitica* (45 Mbp) und damit deutlich größer als das der meisten anderen Zygomyceten. Generell liegt die Genomgröße bei Zygomyceten im Mittel bei 39 Mbp. Zwischen den Genomen von Zygomyceten gibt es starke Unterschiede. Das Genom von *Umbeopsis ramanniana* ist gerade einmal 23 Mbp groß, das Genom von *Mucor heterogamus* hingegen über 54 Mbp. Neben der Integration fremder DNA parasitischen Ursprungs ist auch eine Häufung von Genduplikationen, wie sie bereits aus anderen Zygomyceten, wie zum Beispiel *Rhizopus oryzae* (46 Mbp), bekannt sind (Cornell et al., 2007), eine mögliche Ursache für die Genomgröße von *Absidia glauca*. Ein anderer Faktor, der die Größe eines Genoms beeinflusst, ist die Anzahl mobiler genetischer Elemente. Der GC-Gehalt in Exons und Introns ist in *Absidia glauca* (Exons: 48 %, Introns: 29 %) etwas höher als in *Parasitella parasitica* (Exons: 44 %, Introns: 24 %). Die Anzahl an Genen und tRNAs ist in *Absidia glauca* (15 594 Gene, 232 tRNAs) höher als in *Parasitella parasitica* (14 507 Gene, 161 tRNAs). Die Länge der Gene aus den beiden Genomen ist mit 1 275 bp beziehungsweise 1 226 bp im Durchschnitt vergleichbar. Gene aus *Absidia glauca* beinhalten häufiger Introns (im Durchschnitt 4,5 Introns pro Gen), als Gene aus *Parasitella parasitica* (im Durchschnitt 3,6 Introns pro Gen). Die Länge der Exons ist in *Absidia glauca* daher rund 60 bp kürzer. Die Länge der Introns entspricht mit über 80 bp der in *Parasitella parasitica*.

Sequenzanalysen bestätigen frühere Ergebnisse, die das regelmäßige Auftreten repetitiver DNA mit Hybridisierungsstudien belegen (Burmester und Wöstemeyer, 1994). Der Anteil repetitiver Sequenzen ist in *Absidia glauca* über 25 % höher als bei *Parasitella parasitica*. Vergleicht man die repetitiven Sequenzen beider Genome, zeigt sich, dass *Absidia glauca* zwar einen höheren Anteil repetitiver Sequenzen im Genom trägt, die repetitiven Sequenzen in *Parasitella parasitica* allerdings aus weitaus mehr Familien transponierbarer Elemente stammen und daher vielfältiger in ihrer Erscheinung und Funktion sind. Beide Zygomyceten besitzen transponierbare Elemente aus Familien, für die horizontaler Gentransfer dokumentiert wurde. Die Anzahl dieser Elemente ist bei *Parasitella parasitica* um das fünffache höher als bei *Absidia glauca*. In *Absidia glauca* finden sich 166 transponierbare Elemente aus vier Familien, die es nicht in *Parasitella parasitica* gibt. Bei einer davon ist horizontaler Gentransfer dokumentiert. Umgekehrt besitzt *Parasitella parasitica* 2 737 transponierbare Elemente aus acht Familien, die nicht in *Absidia glauca* zu finden sind. Bei drei davon ist horizontaler Gentransfer dokumentiert. Der Anteil repetitiver Sequenzen, die keiner bekannten Familie zugeordnet werden konnten, ist bei

Parasitella parasitica höher als bei *Absidia glauca*.

Die funktionelle Annotation der Gene aus *Absidia glauca* mit BLAST zeigt eine ungewöhnlich hohe Ähnlichkeit dieser Gene zu Genen aus *Mucor*-artigen Zygomyceten und das, obwohl *Absidia glauca* zur Familie der Cunninghamellaceae gehört. Die Sequenzähnlichkeit beschränkt sich nicht auf Gene aus *Parasitella parasitica*, sondern tritt besonders häufig bei Genen aus *Rhizopus microsporus*, *Mucor circinelloides* und *Mucor ambiguus* auf. Dieser Umstand erschwert die Suche nach horizontalem Gentransfer zwischen Wirt und Parasit. Für über 1 400 Gene stammt der beste BLAST-Treffer aus *Parasitella parasitica*. Für 22 % von ihnen wurde keine Proteindomäne gefunden. Ihre Funktion ist daher unbekannt.

4.1.2 Mitochondriale DNA

Der parasitische Lebensstil von *Parasitella parasitica* führt zur Fusion von Hyphen des Wirts und des Parasiten (Burgeff, 1924). Dieser Prozess ist mit dem Transport von Organellen vom Parasiten hinein in den Wirt verbunden. Auf diese Weise kommt es regelmäßig zur Bildung genetischer Rekombinanten zwischen Wirt und Parasit (Burmester et al., 2013).

In Pilzen stellen die *petite*-Mutanten aus *Saccharomyces cerevisiae* die bekannteste eigennützige (selfish) mitochondriale DNA dar (Williamson, 2002). Die mitochondrialen Genome der *rho*⁻-*petite*-Mutanten enthalten große Deletionen. Sie sind nicht in der Lage zu atmen und bilden deshalb nur kleine Kolonien, indem sie ausschließlich gären. *Saccharomyces cerevisiae* ist eine fakultativ aerobe Hefe. Der teilweise Verlust der mitochondrialen DNA tötet die Zelle daher nicht, sondern führt zu vermindertem Wachstum im Vergleich zum Wildtyp. Beide Formen von Mitochondrien können gleichzeitig in einer Population existieren. Unter diesen Umständen verbreitet sich die defekte mitochondriale DNA bei der Vermehrung unter der meiotischen Nachkommenschaft. Auf der mechanistischen Ebene wird dieser Effekt von einer hoch effizienten Replikation erzeugt, die auf der Existenz multipler Replikationsursprünge beruht (MacAlpine et al., 2001). Von mehrkerigen Ascomyceten weiß man, dass die Toleranz für genetische Heterogenität den Austausch und die Rekombination von Zellkernen zwischen Pilzkolonien ermöglicht. Diese Mechanismen sind ein wichtiger Beitrag für die Evolution und Vielfältigkeit der Pilze (Roper et al., 2011). Im Gegensatz dazu bildet der Wirt *Absidia glauca* keine Anastomosen. Dieser Pilz scheint sogar Mechanismen entwickelt zu haben, um seine Zellkerne vor fremder DNA zu schützen. So können nur kurze Fragmente der DNA von *Parasitella parasitica* permanent in das Genom von *Absidia glauca* integriert werden (Burmester et al., 2013).

In dieser Arbeit wurde die Synthese mitochondrialer DNA von Zygomyceten untersucht, um die Entwicklung der mitochondrialen DNA von Wirt und Parasit nach der Infektion und damit verbundenen Fusion nachzuvollziehen und eventuelle Unterschiede

zu mitochondrialer DNA von Zygomyceten, die nicht von *Parasitella parasitica* infiziert werden, zu identifizieren.

Das erstaunlichste Ergebnis ist die hohe Anzahl an Endonukleasen in der mitochondrialen DNA des Parasiten *Parasitella parasitica*. Der Fusionsparasit besitzt 7 GIY-YIG- und 19 LAGLIDADG-Endonukleasen in seiner mitochondrialen DNA. Der Wirt *Absidia glauca* besitzt ebenfalls eine für Zygomyceten überdurchschnittlich hohe Anzahl an Endonukleasen, 8 GIY-YIG- und 4 LAGLIDADG-Endonukleasen.

Homing-Endonukleasen in Gruppe-I-Introns sind eigennützige (selfish) Gene. Sie sind in der Lage, ihre Position im Genom zu verändern (Stoddard, 2014). Die Sequenzidentität des Zielgens muss hoch genug sein, damit eine Kopie der Homing-Endonuklease an der neuen Position eingebaut werden kann. Die Integration in das Wirtsgenom schädigt das Gen nicht, da sich die Homing-Endonuklease in ein Intron und nicht in den codierenden Bereich des Gens einfügt.

Zu Beginn der Analyse stand die Annahme, dass aufgrund des invasiven Charakters der in Introns codierten Homing-Endonukleasen die Anzahl mobiler genetischer Elemente in potentiellen Wirten höher ist, als in Pilzen, die nicht infiziert werden. Der Vergleich der mitochondrialen DNA verschiedener Zygomyceten unterstützt diese Hypothese nicht. Entweder erfolgt der Transport von mobilen genetischen Elementen unabhängig vom Infektionsprozess oder die Wirte haben einen Mechanismus entwickelt, um ihre mitochondriale DNA vor fremden genetischen Elementen zu schützen.

Es ist möglich, dass in Zygomyceten Mitochondrien generell nur selten fusionieren und DNA kaum transferiert wird. *Mucor circinelloides*, der ein potentieller Wirt für *Parasitella parasitica* ist, besitzt zum Beispiel nur ein Intron und das enthält keine Endonuklease. Im Gegensatz dazu finden sich in der mitochondrialen DNA von *Phycomyces blakesleeanus*, der nicht von *Parasitella parasitica* infiziert wird 12 Endonukleasen. Diese sitzen alle in Introns. Möglicherweise haben diese Endonukleasen ihren Ursprung in einer Ansammlung von Genen eines gemeinsamen Vorfahren dieser Zygomyceten. Die hohe Sequenzidentität zwischen Endonukleasen aus *Absidia glauca* und *Phycomyces blakesleeanus*, die nicht nah miteinander verwandt sind (Hoffmann et al., 2013), unterstützt diese Hypothese. Die Sequenzidentität der Endonukleasen spiegelt nicht die verwandtschaftlichen Beziehungen der Zygomyceten wider. Ein Neighbor-Joining-Baum aller betrachteten Endonukleasen veranschaulicht diese Besonderheit (Abbildung 4.1).

Sequenzen mit höherer Identität zu den Endonukleasen eines Zygomyceten fanden sich oft in mehreren nur entfernt verwandten Pilzen. Die größte Ähnlichkeit für Endonukleasen aus *Absidia glauca* hatten Sequenzen aus *Backusella circina* (89 %), *Phycomyces blakesleeanus* (84 %, 97 %) und *Rhizopus oryzae* (84 %). Endonukleasen aus *Absidia glauca*, die am ähnlichsten zu Sequenzen aus *Parasitella parasitica* waren, ergaben geringere Sequenzidentitäten (56 %, 63 %). 66 % der verglichenen Endonukleasen befanden sich innerhalb des gleichen Gens. Die meisten lagen innerhalb von *cob* und *coxI*.

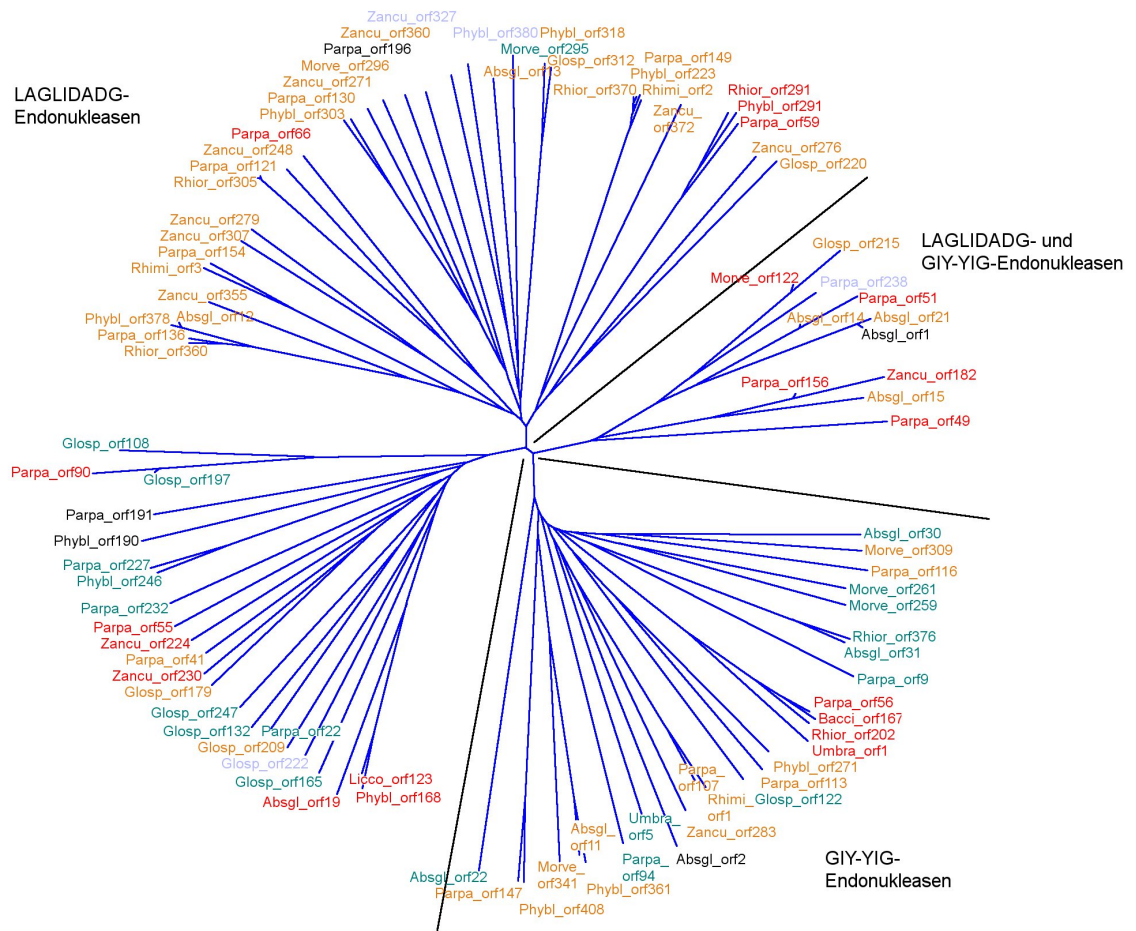


Abbildung 4.1: Neighbor-Joining-Baum der Endonukleasen aus Zygomyceten.

Die Sequenzähnlichkeit der Endonukleasen spiegelt nicht die phylogenetische Verwandtschaft der Zygomyceten wider.

(Abbildung nach Ellenberger, Burmester und Wöstemeyer (2016b))

Innerhalb des Neighbor-Joining-Baums bildeten diese Endonukleasen keine eindeutig abgegrenzten Teilbäume aus. Die größte zusammenhängende Gruppe bildete sich aus LAGLIDADG-Endonukleasen, die sich in *coxI* befanden. Endonukleasen aus *cob* bildeten Gruppen aus drei oder vier Sequenzen oder standen allein im Baum. Endonukleasen, die nicht aus Introns, sondern aus Bereichen zwischen zwei Genen stammen, waren über den gesamten Baum verteilt.

Unter Berücksichtigung der Rekombinationsprozesse in Zellkernen auf der einen Seite und in Mitochondrien auf der anderen Seite lässt sich schlussfolgern, dass sich die Prinzipien, die der Rekombination beider Arten von DNA zugrunde liegen, unterscheiden. Die Rekombination von Genen im Zellkern von Wirt und Parasit wurde für das genetische System aus *Absidia glauca* und *Parasitella parasitica* auf experimenteller Ebene nachgewiesen und ist eher die Regel als die Ausnahme. Die Rekombination von Mitochondrien konnte dagegen weder im Experiment, noch in dieser Arbeit nachgewiesen werden. Mitochondrien sind in Zygomyceten nicht primär das Ergebnis von Rekombinationsereignissen. Die Mobilität genetischer Elemente mit Hilfe von Homing-Endonukleasen ist in Zygomyceten von Interaktionen zwischen Wirt und Parasit unabhängig. *Parasitella parasitica* ist ein Reservoir für Endonukleasen, während zum Beispiel *Mucor circinelloides* keine Endonukleasen besitzt, obwohl beide Pilze zu den *Mucoraceae* gehören. Die Anzahl an Endonukleasen scheint daher über Spezies-spezifische Kontrollmechanismen der mitochondrialen DNA gesteuert zu werden.

4.2 Posttranslationale Regulation innerhalb der Trisporsäure-Synthese

4.2.1 Posttranslationale Regulation von TSP1

Die Suche nach TSP1 in Zygomyceten führte überraschenderweise zur Identifizierung einer Vielzahl von TSP1- und TSP1-ähnlichen Proteinen. Bei TSP1 handelt es sich um ein für den Trisporsäure-Syntheseweg des (–)-Kreuzungstypen spezifisches Enzym. Eine erste Annahme, nach der sich unter den neu gefundenen Proteinen die noch unbekannten Enzyme des (+)-Kreuzungstypen befinden, hat sich nicht bestätigt. Die betrachteten (+)-Kreuzungstypen von *Phycomyces blakesleeana*, *Rhizopus oryzae* und *Thamnidium elegans* besitzen nur ein TSP1-Protein. Dieses hat in allen drei Pilzen eine hohe Sequenzähnlichkeit mit dem Original-TSP1-Protein aus *Mucor mucedo* (Czempinski et al., 1996).

Die Anzahl verschiedener TSP1-Varianten unterscheidet sich deutlich bei den betrachteten Zygomyceten und schwankt zwischen eins und neun. Da für *Mucor mucedo* und *Rhizomucor pusillus* die vollständigen Genome nicht vorliegen, können weitere TSP1-Varianten in diesen beiden Zygomyceten nicht ausgeschlossen werden. In *Rhizomucor pusillus* existiert eine zweite, als Xylose-Reduktase annotierte Sequenz (UniProt: I7HD84),

die eine Sequenzidentität von 98 % mit einem TSP1-Protein aus *Mucor circinelloides* 1006PhL (JGI: 511006243) liefert. Der Sequenzname des Proteins ist das Ergebnis des Annotationsprozesses und wurde nicht experimentell überprüft. Xylose-Reduktasen wandeln Xylose in Xylitol um und vollziehen den ersten Schritt des Xylose-Abbaus. Der Xylose-Abbau in Zygomyceten wurde an *Zygorhynchus moelleri* in einem Experiment getestet (Emilia Schulz, Lehrstuhl für Allgemeine Mikrobiologie und Mikrobengenetik, FSU Jena, persönliche Kommunikation). Das Myzel wurde auf einer Petrischale, die als Kohlenstoffquelle Xylose und als Färbereagenz Kongorot enthielt, kultiviert. Baut der Pilz Xylose enzymatisch ab, verschwindet die Färbung, sodass um den wachsenden Pilz ein weniger gefärbter Hof entsteht. Dieser Effekt konnte bei *Zygorhynchus moelleri* nicht beobachtet werden. Xylose-Abbau ist für Zygomyceten nicht dokumentiert. Die Xylose-Reduktase aus *Rhizomucor pusillus* wurde daher als zweites TSP1-Protein berücksichtigt.

In den homothallischen Zygomyceten *Mucor heterogamus*, *Mycotypha africana* und *Syncephalastrum racemosum* wurden zwei, drei, beziehungsweise sechs TSP1-Proteine identifiziert. Es lies sich keine Präferenz für eine bestimmte Anzahl an TSP1-Proteinen feststellen.

Die durchschnittliche Sequenzidentität aller TSP1-Proteine betrug 56 %. Die Sequenzidentität war höher für TSP1-Sequenzen einer Variante aus unterschiedlichen Pilzen, als für verschiedene TSP1-Varianten aus einem Organismus. Die hohe Sequenz- und Strukturähnlichkeit unterstützt die These, dass horizontaler Gentransfer die Ursache der Akkumulation von Genen ist, die TSP1-Proteine codieren. Es ist möglich, dass Zygomyceten mit Hilfe unterschiedlicher TSP1-Varianten in der Lage sind, ihr Substratangebot zu erweitern und unterschiedliche Trisporoid-Serien parallel umzusetzen.

Parasitella parasitica besitzt sechs mehr oder weniger verschiedene Gene für TSP1- und TSP1-ähnliche Proteine. Eine Erklärung für das unerwartete Auftreten multipler TSP1-Varianten ist die evolutionäre Anpassung an unterschiedliche Substrate. Auf diese Weise ist der Pilz in der Lage, mit verschiedenen anderen Pilzen, besonders mit potentiellen Wirten, zu kommunizieren, obwohl es sich bei deren Pheromonen unter Umständen um Substanzen handelt, die nicht primär für die innerartliche, sexuelle Kommunikation benutzt werden. Dieser hypothetische posttranslationale Regulationsmechanismus wurde *in silico* mit der Modellierung von TSP1-Dimeren getestet, um mögliche Schaltstellen zwischen parasitischer und sexueller Kommunikation zu erkennen. Experimentelle genetische Analysen werden von speziellen organismischen Eigenschaften bei Zygomyceten erschwert. Die hohe Sequenzähnlichkeit in bestimmten Bereichen sorgt dafür, dass man bis jetzt von nur einem einzigen Gen für TSP1-Proteine ausgegangen ist. Experimentelle Studien in *Mucor mucedo* deutenden bereits auf die Bildung von enzymatisch aktiven Dimeren hin (Schimek et al., 2005).

Das Trisporinsäure-Niveau von *Parasitella parasitica* genügt nicht, um die Entwicklung sexueller Strukturen zu induzieren. Mindestens ein Partner muss auf einem Wirt wach-

sen, um Zygosporien bilden zu können (Burgeff, 1924). Die Analyse der TSP1-Dimere zeigt, dass der Parasit die Trisporinsäure-Produktion des Wirts anregt, um das gesteigerte Pheromon-Niveau für die eigene sexuelle Entwicklung zu nutzen. *Absidia glauca* wird von einer Infektion mit *Parasitella parasitica* nicht nachhaltig geschädigt. Die TSP1-Proteine aus *Absidia glauca* haben sich als erstaunlich robust gegenüber der Bindung an andere TSP1-Proteine erwiesen. Sowohl die Dimerisierung mit anderen TSP1-Proteinen aus *Absidia glauca*, als auch die Bindung an TSP1-Proteine aus *Parasitella parasitica* führte nur zu einer geringen Abnahme der enzymatischen Aktivität. Inaktivierung, wie sie bei CEP13671 zu sehen ist, trat nur bei Dimeren auf, die eine Kombination beider Original-TSP1-Sequenzen aus *Absidia glauca* und *Parasitella parasitica* darstellen.

Zwei spezialisierte TSP1-Proteine vermitteln die sexuelle und parasitische Kommunikation in *Parasitella parasitica*. Das TSP1-Protein CEP13671 hat die größte Ähnlichkeit zu dem Original-TSP1-Protein aus *Mucor mucedo* (Czempinski et al., 1996) und kommt bei sexuellen Interaktionen, genau wie bei anderen Zygomyceten, zum Einsatz. Das TSP1-Protein CEP10423 ist spezifisch für parasitische Interaktionen und bildet mit den TSP1-Proteinen aus *Absidia glauca* enzymatisch aktive Dimere. Für die Inaktivierung von TSP1-Dimeren in *Parasitella parasitica* ist eine flexible Schleifenregion verantwortlich. Diese schiebt sich in die Trisporoid-Bindetasche und verhindert so, dass ein Trisporoid mit dem aktiven Zentrum des Enzyms in Kontakt kommt. Eine Reaktion ist auf diese Weise nicht mehr möglich. Die Stabilität der inaktiven Homodimere wurde mit der Modellierung von *in silico* Mutanten untersucht. Modifikationen in der flexiblen Schleifenregion führten zu unterschiedlicher enzymatischer Aktivität. Erstaunlicherweise genügte bereits der Austausch von drei Aminosäuren (Y132E, V133E, H135E) innerhalb der internen Schleifenregion, um aktive Homodimere entstehen zu lassen. Diese ähneln dem aktiven Homodimer aus dem TSP1-Protein CEP41023. Bei enzymatisch aktiven Dimeren ist die $\alpha 5$ -Helix Teil der Dimerisierungsschnittstelle, genau wie bei der Xylose-Reduktase aus *Candida tenuis* (Kavanagh et al., 2002).

4.2.2 Posttranslationale Regulation von TSP2

Die Suche nach TSP2 in Zygomyceten führte ebenfalls zur Identifizierung mehrerer TSP2- und TSP2-ähnlicher Proteine. Bei TSP2 handelt es sich genau wie bei TSP1 um ein für den Trisporinsäure-Syntheseweg des (–)-Kreuzungstypen spezifisches Enzym. 4-Dihydrotrisorin-Dehydrogenase (TSP2) ist eine kurzkettige Dehydrogenase. Man unterteilt diese Proteine in fünf Familien. Dabei unterscheidet man klassische, erweiterte, intermediäre, divergente und komplexe kurzkettige Dehydrogenasen (Kallberg et al., 2002). Die Familien werden von unterschiedlichen Motiven im aktiven Zentrum und der Region für die Cosubstrat-Bindung definiert. Die Länge der Sequenzen variiert zwischen den Familien. Auch die Cosubstrat-Spezifität ist unterschiedlich. TSP2 zeigt die typischen Motive einer klassischen kurzkettigen Dehydrogenase in der Cosubstrat-Binderegion (TGxxxGxG)

und dem aktiven Zentrum (YxxxK). Um verschiedene Unterfamilien zu unterscheiden, muss man sich die β A, α B und β B Abschnitte der Proteinsequenzen genauer anschauen (Kallberg et al., 2002). Die Nummerierung in Klammern bezieht sich auf die Nummerierung der $3\alpha,20\beta$ -Hydroxysteroid-Dehydrogenase (PDB: 2HSD). Bei klassischen kurzkettigen Dehydrogenasen, die NADP bevorzugen, kompensieren ein oder zwei positiv geladene Aminosäurereste die zwei negativen Ladungen der 2'-Phosphatgruppe des Cosubstrats. Die Unterfamilien unterscheiden sich anhand der Positionen dieser Aminosäurereste. Sie werden als cP1 (Lys/Arg-15), cP2 (Lys/Arg-37) und cP3 (Lys/Arg-15 und Lys/Arg-37) bezeichnet. Dehydrogenasen, die als Cosubstrat NAD bevorzugen, werden von einer Asparaginsäure oder Glutaminsäure am Ende des zweiten β -Strangs charakterisiert. Diese Unterfamilien heißen cD1d (Asp-36) und cD1e (Glu-36). Wenn es eine negativ geladene Aminosäure an der ersten oder zweiten Position nach dem zweiten β -Strang gibt, wird das Protein der Unterfamilie cD2 (Asp/Glu-37) oder cD3 (Asp/Glu-38) zugeordnet. TSP2 gehört in die Unterfamilie cP3 der klassischen kurzkettigen Dehydrogenasen, weil es zwei Arginine an erster (Arg-12) und zweiter (Arg-34) Position besitzt. Das bestätigt frühere Aussagen über die Cosubstrat-Präferenz von TSP2-Proteinen (Wetzel et al., 2009).

Für Proteine, die eine hohe Sequenzidentität von bis zu 50 % mit TSP2-Proteinen aufweisen, sind die dreidimensionalen Proteinstrukturen unbekannt. Sie kommen deshalb nicht für die Modellierung von TSP2-Strukturen in Frage. Die für die Strukturvorhersage verwendeten Proteine haben eine Sequenzidentität von rund 20 %. Diese Übereinstimmung ist geringer als bei TSP1 (46 %) und ist für Homologie-basierte Strukturmodellierung kritisch. Generell liegt die Sequenzidentität für paarweise Vergleiche von kurzkettigen Dehydrogenasen zwischen 15 % und 30 % (Oppermann et al., 2003). Die Grauzone, ab der eine Sequenzähnlichkeit zu gering ist, um von ihr auf die zugrundeliegende Proteinstruktur zu schließen, liegt im Allgemeinen bei 30 % (Khor et al., 2015). Da die Kernstruktur bei TSP2-Proteinen und kurzkettigen Dehydrogenasen gut konserviert ist, führt die Homologie-Modellierung trotz niedriger Sequenzidentitäten zwischen Vorlage und Modell zu guten Strukturvorhersagen.

In den Genomen von *Absidia glauca* und *Parasitella parasitica* wurden je zwei TSP2-Proteine identifiziert. Andere Zygomyceten besitzen bis zu sieben TSP2-Proteine. Wie bei TSP1 gibt es auch bei TSP2 keine Präferenz für eine bestimmte Anzahl an Sequenzen abhängig davon, ob ein Zygomycet homothallisch oder heterothallisch, Wirt oder Parasit ist.

Kurzkettige Dehydrogenasen bilden Dimere und Tetramere (Oppermann et al., 2003). Diese Tetramere besitzen zwei unterschiedliche Schnittstellen für die Bindung an einen Partner. Man unterscheidet drei orthogonale molekulare Achsen (P, Q und R), die für die Dimerisierung wichtig sind (Rossmann et al., 1973). Der engste Kontakt entsteht an der Q-Achse über die beiden Helices α E und α F. Das TSP2-Protein CEP13616 bildet

typische Dimere über diese beiden Helices. Es gibt zwei Carbonyl-Reduktasen, die zur Gruppe der kurzkettigen Dehydrogenasen gehören und deren Proteinstrukturen als Monomere charakterisiert wurden (PDB: 1N5D, 1WMA) (Ghosh et al., 2001; Tanaka et al., 2005). Diese beiden Proteinstrukturen gehören zu den für die Homologie-Modellierung der TSP2-Proteine verwendeten Vorlagenstrukturen. Das zweite TSP2-Protein CEP13991 bildet nicht diese für kurzkettige Dehydrogenasen typischen Dimere. Es ist möglich, dass es sich bei diesem Enzym um ein als Monomer auftretendes Enzym handelt. Eine andere Erklärung für diese untypische Dimerbildung könnte in dem Einfluss liegen, denn die zusätzlichen Bindungspartner in einem Tetramer auf die Dimerisierungszonen ausüben. Das Protein-Protein-Docking mit ClusPro berücksichtigt nur die Wechselwirkungen zwischen zwei Bindungspartnern.

4.2.3 Interaktion zwischen TSP1 und TSP2

Es wurde gezeigt, dass in *Parasitella parasitica* die enzymatische Aktivität von TSP1-Proteinen posttranslational über die Bildung von Dimeren mit anderen TSP1-Proteinen reguliert wird. Ob auch TSP2-Proteine als regulatorische Bindungspartner in Frage kommen, wurde anhand von Protein-Protein-Docking mit beiden Enzymen aus *Parasitella parasitica* getestet. Das Ergebnis waren Dimere mit verminderter enzymatischer Aktivität. Der Mechanismus der Inaktivierung unterscheidet sich von dem in TSP1-Dimeren. Bei Dimeren mit TSP2 wird das aktive Zentrum der Proteine nicht von dem Partnerprotein verschlossen. Es kommt zu einer Stauchung der Proteinstruktur, die zu einer Verengung der Trisporoid-Bindetasche führt. TSP2-Proteine binden an einer anderen Stelle des TSP1-Proteins als TSP1-Proteine.

Die Modellierung der Bindung von TSP1-Proteinen an TSP2-Proteine führte für die beiden TSP2-Varianten aus *Parasitella parasitica* zu unterschiedlichen Ergebnissen. Während CEP13616 eine hohe enzymatische Aktivität zeigte, war CEP13991 nur in einem Dimer aktiv. Diese Ergebnisse zeigen, dass es sich auch bei den beiden TSP2-Proteinen um spezialisierte Enzyme handelt, die nach der Bindung an andere Proteine eine unterschiedliche enzymatische Aktivität aufweisen.

4.3 Detektion von horizontalem Gentransfer zwischen Protozoen und Bakterien

Im Forschungsalltag werden in der Regel Organismen untersucht, deren Genome noch nicht vollständig sequenziert wurden. Aus diesem Grund sind Programme, die mit ganzen Genomen arbeiten, auf diese Aufgaben nicht anwendbar. Oft sind die genauen verwandtschaftlichen Beziehungen der zu betrachtenden Organismen unklar. Gerade bei Protozoen wird immer noch über deren Taxonomie diskutiert. Das Interesse für die Entwicklung

eines bestimmten Gens erfordert nicht immer genaue Kenntnisse über die Entwicklung einer Spezies. Trotzdem werden Daten von anderen Organismen benötigt, um einen Speziesbaum zu erzeugen, mit dem ein Genbaum verglichen werden kann, um ungewöhnliche Ereignisse zu ermitteln. In dieser Arbeit wurde der Frage nachgegangen, ob es möglich ist, einzelne HGT-Ereignisse zwischen Prokaryoten und Eukaryoten in kleinen Datensätzen gezielt zu identifizieren, ohne komplette Genome oder einen Speziesbaum zur Verfügung zu haben.

4.3.1 Identifizierung von horizontalem Gentransfer

Die Unterscheidung von endosymbiotischem (EGT) und horizontalem Gentransfer (HGT) zwischen Prokaryoten und Eukaryoten mit Hilfe phylogenetischer Methoden ist nicht immer einfach. Da nur ein kleiner Teil der prokaryotischen Vielfalt bereits sequenziert wurde, ist es natürlich möglich, dass die Spezies, aus der das transferierte Gen stammt, noch gar nicht in einer Datenbank zu finden ist. Deshalb erscheinen HGT-Ereignisse oft älter, als sie in Wirklichkeit sind. Gene, die das Ergebnis rezenter HGT-Ereignisse zwischen Organellen und Zellkern sind, werden dagegen oft gemeinsam mit dem Spendergenom sequenziert und sind leichter zu identifizieren.

Es gibt Unterschiede bei der Suche nach horizontalem Gentransfer, je nachdem ob es sich um einen Transfer innerhalb von Bakterien oder von Bakterien zu Protozoen handelt. Auch zwischen den einzelnen Detektionsmethoden gibt es Unterschiede. Daher werden in dem in dieser Arbeit vorgestellten Algorithmus verschiedene Baum-basierte und nicht Baum-basierte Methoden kombiniert. Das erhöht die Sensitivität und den Informationsgehalt der Suche. Der neue Algorithmus ist in der Lage, mit kleinen Datensätzen zu arbeiten und den Anteil der phylogenetischen Analyse auf kleine Neighbor-Joining-Bäume zu reduzieren, die nur als eine Art interne Datenstruktur verwendet werden und nicht der detaillierten Aufklärung phylogenetischer Verwandtschaften dienen. Zusätzlich liefern die Ergebnisse der BLAST-Suche erste Hinweise auf mögliche HGT-Ereignisse und den putativen Ursprungsorganismus des eingewanderten genetischen Materials. Die Aufgabe dieses Programms ist es nicht, die exakte Topologie eines phylogenetischen Baumes zu ermitteln. Das lässt sich mit anderen Ansätzen erreichen. Zum Beispiel mit dem Vergleich von ITS-Regionen oder kombinierten Sequenzen bestehend aus den beiden Untereinheiten ribosomaler DNA, Aktin und eines Elongationsfaktors. Die vorgestellte Methode soll eine Sequenz lediglich eindeutig einem eukaryotischen oder prokaryotischen Teilbaum zuordnen. Wenn eine eukaryotische Sequenz innerhalb eines prokaryotischen Teilbaums auftaucht und der Abstand zwischen dem eukaryotischen und prokaryotischen Teilbaum ausreichend ist, liegt ein HGT-Ereignis vor.

HGT Calculator (siehe Abschnitt 3.6) wurde an Genen aus Trypanosomatiden getestet. Alves et al. (2013) fanden deutliche Unterschiede im Nährstoffbedarf von Trypanosomatiden, die Endosymbionten haben und solchen ohne Endosymbionten. Der Algorithmus

wurde auf Gene von Trypanosomatiden mit unterschiedlichem Lebensstil angewandt, inklusive einem für Menschen nicht pathogenen *Trypanosoma*-Stamm, *Trypanosoma rangeli* (Stoco et al., 2014).

4.3.2 Vorteile gezielter Identifizierung von horizontalem Gentransfer

Alle Ergebnisse beweisen die Funktionalität des neuen Ansatzes zum Auffinden von HGT-Ereignissen zwischen Bakterien und Protozoen. HGT Calculator ist in der Lage gezielt einzelne Sequenzen auszuwerten, ohne auf vollständige Genome oder umfangreiche phylogenetische Methoden angewiesen zu sein. Gerade die Erstellung eines gut belegbaren Spezies-Baums ist unter Umständen schwierig. Zumal das Konzept „Spezies“ selbst in Protozoen nicht eindeutig ist (Schlegel und Meisterfeld, 2003). Die Anwendung verschiedener phylogenetischer Methoden kann zudem zu unterschiedlichen Spezies-Bäumen führen (Leaché und Rannala, 2010). HGT Calculator ermöglicht es, über die Wahrscheinlichkeit eines HGT-Ereignisses einzelner Gene zu entscheiden. Damit richtet sich diese Methode nach der experimentellen Realität, die die meisten HGT- und Phylogenie-Projekte beschränkt.

Gene für Aconitat-Hydratase, Adenosylhomocysteinase, Enolase, Guanylat-Kinase, S-Adenosylmethionin-Synthase, Nukleosid-Diphosphat-Kinase, Ribonukleosid-Diphosphat-Reduktase, Serin-Hydroxymethyltransferase, Triosephosphat-Isomerase und Thymidylat-Synthase wurden als Negativkontrolle für den horizontalen Gentransfer in Trypanosomatiden (Alves et al., 2013) eingesetzt. Die HGT-Werte sind bei allen Genen dieses Datensatzes null (Tabelle 2.6). Ursprüngliche Sequenzen werden sowohl von der Auswertung der BLAST-Ergebnisse, als auch von der Multi-Level-Analyse einfacher Neighbor-Joining-Bäume als HGT-Ereignis zurückgewiesen.

Das Auffinden von HGT-Ereignissen ist und bleibt ein aktuelles Thema innerhalb der Phylogenetik. In Bakterien hat sich gezeigt, dass HGT-Ereignisse zur Ausbildung von Antibiotikaresistenzen, Pathogenität oder neuer metabolischer Eigenschaften beigetragen haben (Juhas et al., 2009). Inzwischen weiß man, dass HGT-Ereignisse auch in Eukaryoten eine entscheidende Rolle spielen und das über die evolutionäre Entstehung von Mitochondrien und Plastiden hinaus. Eine aktuelle Studie beschreibt sogar einen möglichen Gentransfer von marinen Bakterien zu Fischen (Sun et al., 2015). Die Anzahl an publizierten HGT-Ereignissen zwischen Prokaryoten ist noch immer deutlich größer als bei Eukaryoten (Andersson, 2005), obwohl horizontaler Gentransfer inzwischen auch bei Pilzen (Fitzpatrick, 2012), Würmern (Boto, 2014) und Protozoen (Huang, Mullapudi, Lancto et al., 2004) nachgewiesen wurde. In Protisten, die phagotroph leben, tritt horizontaler Gentransfer mit einer ähnlichen Häufigkeit auf, wie bei Prokaryoten (Andersson, 2005). Solche HGT-Ereignisse sind sogar von medizinischem Interesse. Eine Arbeit über HGT-Ereignisse in dem weltweit verbreiteten Protozoon *Cryptosporidium parvum* (Stripen et al., 2004) zeigt den Zusammenhang zwischen Schwierigkeiten bei der Behandlung

von Cryptosporidiose und horizontalem Gentransfer.

Es wurden 20 Gene analysiert, die aus Trypanosomatiden mit unterschiedlicher Lebensweise stammen. Verglichen wurden Organismen mit und ohne Endosymbionten, von denen einige pathogen sind. Die Gene der Isocitrat-Dehydrogenase aus *Leishmania major* und *Strigomonas culicis* dienen als anschauliches Beispiel für diesen neuen Algorithmus zum Auffinden von HGT-Ereignissen zwischen Bakterien und Protozoen. Isocitrat-Dehydrogenase wird von Opperdoes und Michels (2007) als HGT-Ereignis in Trypanosomatiden beschrieben. In dieser Veröffentlichung existiert das mit horizontalem Gentransfer entstandene Gen nur in *Leishmania major* (LmjF33.2550) und nicht in *Trypanosoma brucei* und *Trypanosoma cruzi*, obwohl es sich hierbei um Trypanosomatiden handelt, die mit *Leishmania major* verwandt sind. Mit dem neuen Ansatz des HGT Calculators wurde das korrespondierende Gen in einem noch nicht auf horizontalen Gentransfer untersuchten Organismus identifiziert und Isocitrat-Dehydrogenase als *Leishmania*-spezifisches HGT-Ereignis bestätigt. In *Trypanosoma rangeli* existiert das HGT-Gen ebenfalls nicht. Die Isocitrat-Dehydrogenase aus *Trypanosoma rangeli* (UniProt: A0A061IWC2) ergab einen HGT-Wert von null und ist eindeutig kein HGT-Ereignis.

Der neue Algorithmus wurde an weiteren Genen getestet, die in den Publikationen von Opperdoes und Michels (2007) und Alves et al. (2013) als HGT-Ereignisse in *Leishmania major* und *Strigomonas culicis* beschrieben sind. HGT Calculator identifizierte die zehn betrachteten Gene eindeutig als HGT-Ereignisse und unterscheidet dabei sogar zwischen einmalig stattgefundenen, mehrfach auftretenden und lang zurückliegenden HGT-Ereignissen. Auf diese Weise ließen sich für einen bislang noch nicht auf HGT-Ereignisse untersuchten Organismus, *Trypanosoma rangeli*, erste Aussagen über mit bereits veröffentlichten HGT-Genen korrespondierenden Genen treffen.

Wie spezifisch HGT Calculator arbeitet, zeigt sich am deutlichsten am Beispiel der Homoserin-Dehydrogenase in Trypanosomatiden. Sowohl in *Leishmania major* als auch in *Strigomonas culicis* liegt der Ursprung der Gene für Homoserin-Dehydrogenase in HGT-Ereignissen (Alves et al., 2013). Das Gen existiert in Trypanosomen mit und ohne Endosymbionten. HGT Calculator identifizierte die Gene aus *Leishmania major* und *Strigomonas culicis* eindeutig als HGT-Ereignisse. Der Algorithmus unterscheidet dabei zwei unterschiedliche HGT-Ereignisse. Der Transfer von einem Endosymbionten zu seinem Wirt ergab den höchsten HGT-Wert. Der HGT-Wert für das HGT-Ereignis in *Strigomonas culicis* ergab einen deutlich höheren HGT-Wert (65), als das HGT-Ereignis in *Leishmania major* (30). Mit dem Auftreten von zwei voneinander unabhängigen HGT-Ereignissen sinkt der HGT-Wert für beide. Für den Transfer in den gemeinsamen Vorfahren von Trypanosomatiden mit und ohne Endosymbionten war der HGT-Wert sogar null. Hier lies sich das HGT-Ereignis erst mit der Auswertung der Zwischenwerte identifizieren. Die HGT-Zwischenwerte waren 0 (30) t, beziehungsweise 0 (40) t. Besonders der hohe Alien-Index bewies, dass es sich bei beiden Sequenzen um HGT-Ereignisse handelt. Schaut man

sich die BLAST-Ergebnisse genauer an, erkennt man, dass die beiden HGT-Ereignisse einen unterschiedlichen Ursprung haben. Die Homoserin-Dehydrogenase aus *Leishmania major* stammt aus einem multiplen HGT-Ereignis zwischen einem Firmicuten und Trypanosomatiden. *Strigomonas culicis* besitzt zwei Gene für Homoserin-Dehydrogenase. Eines ähnelt der Sequenz aus *Leishmania major*, das andere hat seinen Ursprung in einem Endosymbionten. Der beste prokaryotische BLAST-Treffer für das Gen aus *Leishmania major* stammte aus *Lysinibacillus*, einem Firmicuten und wies eine Sequenzidentität von 50 % auf. Der beste prokaryotische BLAST-Treffer für das erste Gen aus *Strigomonas culicis* (UniProt: S9UCM8) stammte ebenfalls aus *Lysinibacillus* mit einer vergleichbaren Sequenzidentität von 54 %. Das zweite Gen (UniProt: S9UM95) hingegen hatte seinen besten BLAST-Treffer in *Candidatus Kinetoplastibakterium blastocrithidii*, einem β -Proteobakterium und Endosymbionten von *Strigomonas culicis*, mit einer Sequenzidentität von fast 100 %. Diese Sequenz war die einzige eukaryotische Sequenz in der Liste der BLAST-Ergebnisse. Für das Gen aus *Leishmania major* und das zweite Gen aus *Strigomonas culicis* fanden sich mehrere andere eukaryotische BLAST-Treffer.

HGT Calculator ermöglicht es, Aussagen über die Wahrscheinlichkeit für horizontalen Gentransfer bei einzelnen Genen zu treffen. Es werden dabei einzelne Sequenzen in kleinen Datensätzen untersucht. Das Programm ist nicht für das Durchsuchen vollständiger Genome gedacht, sondern für eine gezielte Gen-spezifische Anwendung. Dieser neue Ansatz unterscheidet sich von bereits etablierten HGT-Identifikationsalgorithmen wie zum Beispiel RIATA-HGT (Nakhleh et al., 2005) oder TREX (Boc et al., 2012) dadurch, dass er ohne vollständige Genome und einen im Vorfeld zu erstellenden Speziesbaum auskommt. Viele Baum-basierte Anwendungen, wie Pyphy (Sicheritz-Pontén und Andersson, 2001), PhyloGenie (Frickey und Lupas, 2004), DarkHorse (Podell und Gaasterland, 2007) oder HGTector (Zhu et al., 2014) wurden für die Suche nach HGT-Ereignissen zwischen Prokaryoten optimiert und lassen sich nicht ohne weiteres auf Eukaryoten anwenden. In einer aktuellen Publikation (Nguyen et al., 2015) werden solche Verfahren entweder als schwierig zu installieren oder arbeitsintensiv in der manuellen Bedienung und nachträglichen Auswertung beschrieben. HGT Calculator ist eine ausführbare JAR-Datei und benötigt keine echte Installation. Der einzige etwas aufwendigere Schritt ist die Auswahl geeigneter Sequenzen für den Datensatz. Dieser ist allerdings bei allen Anwendungen notwendig und wird mit dem HGT Downloader vereinfacht.

Das Problem der phylogenetischen Rekonstruktion dreht sich für gewöhnlich im Kreis. Die Erstellung eines Speziesbaums erfordert Kenntnisse über die Evolution von Genfamilien, wie zum Beispiel die Duplikation oder der Verlust von Genen. Beides, sowohl die Rekonstruktion eines Genbaums, als auch die Identifikation von Duplikationsereignissen oder Genverlusten benötigt einen gut belegten Speziesbaum (Boussau et al., 2013). Die Reduktion der phylogenetischen Analyse auf die Verwendung als grundlegende interne Datenstruktur ermöglicht es, Gruppierungen von eukaryotischen und prokaryoti-

schen Sequenzen zu identifizieren und umgeht die Notwendigkeit komplexer Maximum-Likelihood- oder wahrscheinlichkeitstheoretischer (Bayesscher) Methoden. Um ausschließlich ungewöhnliche Sequenzverteilungen zu ermitteln, sind ungewurzelte Neighbor-Joining-Bäume ausreichend.

5 Literatur

- Abagyan, R., Totrov, M. und Kuznetsov, D. (1994). ICM - A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J Comput Chem* 15, 488–506.
- Aguileta, G., de Vienne, D.M., Ross, O.N., Hood, E.M., Giraud, T., Petit, E. und Gabaldón, T. (2014). High variability of mitochondrial gene order among fungi. *Genome Biol Evol* 6, 451–465.
- Aldib, I., Gelbcke, M., Soubhye, J., Prévost, M., Furtmüller, P.G., Obinger, C., Elfving, B., Alard, I.C., Roos, G., Delporte, C., Berger, G., Dufour, D., Boudjeltia, K.Z., Nève, J., Dufrasne, F. und Antwerpen, P. Van (2016). Novel bis-arylalkylamines as myeloperoxidase inhibitors: Design, synthesis, and structure-activity relationship study. *Eur J Med Chem* 123, 746–762.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. und Lipman, D.J. (1990). Basic local alignment search tool. *J Mol Biol* 215, 403–410.
- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. und Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25, 3389–3402.
- Alves, J.M.P., Klein, C.C., da Silva, F.M., Costa-Martins, A.G., Serrano, M.G., Buck, G.A., Vasconcelos, A.T.R., Sagot, M.-F., Teixeira, M.M.G., Motta, M.C.M. und Camargo, E.P. (2013). Endosymbiosis in trypanosomatids: the genomic cooperation between bacterium and host in the synthesis of essential amino acids is heavily influenced by multiple horizontal gene transfers. *BMC Evol Biol* 13, 190.
- Andersson, J.O. (2005). Lateral gene transfer in eukaryotes. *Cell Mol Life Sci* 62, 1182–1197.
- Anfinsen, C.B. (1973). Principles that govern the folding of protein chains. *Science* 181, 223–230.
- Archibald, J.M. (2015). Endosymbiosis and eukaryotic cell evolution. *Curr Biol* 25, R911–R921.
- Atilgan, A.R., Durell, S.R., Jernigan, R.L., Demirel, M.C., Keskin, O. und Bahar, I. (2001). Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys J* 80, 505–515.
- Baran, R.H. und Ko, H. (2008). Detecting horizontally transferred and essential genes based on dinucleotide relative abundance. *DNA Res* 15, 267–276.
- Beaudet, D., Terrat, Y., Halary, S., de la Providencia, I.E. und Hijri, M. (2013). Mitochondrial genome rearrangements in *Glomus* species triggered by homologous recombination between distinct mtDNA haplotypes. *Genome Biol Evol* 5, 1628–1643.
- Bergthorsson, U., Adams, K.L., Thomason, B. und Palmer, J.D. (2003). Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* 424, 197–201.

- Bhattacharya, D., Cao, R. und Cheng, J. (2016). UniCon3D: *de novo* protein structure prediction using united-residue conformational search via stepwise, probabilistic sampling. *Bioinformatics* 32, 2791–2799.
- Blakeslee, A.F. (1904). Zygosporangium formation a sexual process. *Science* 19, 864–866.
- Boc, A., Diallo, A.B. und Makarenkov, V. (2012). T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic Acids Res* 40, W573–W579.
- Bolger, A.M., Lohse, M. und Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.
- Boto, L. (2014). Horizontal gene transfer in the acquisition of novel traits by metazoans. *Proc R Soc* 281, 20132450.
- Boussau, B., Szöllősi, G.J., Duret, L. Laurent, Gouy, M., Tannier, E. und Daubin, V. (2013). Genome-scale coestimation of species and gene trees. *Genome Res* 23, 323–330.
- Bu'Lock, J.D., Drake, D. und Winstanley, D.J. (1972). Specificity and transformations of the trisporic acid series of fungal sex hormones. *Phytochemistry* 11, 2011–2018.
- Burgeff, H. (1924). Untersuchungen über Sexualität und Parasitismus bei Mucorineen. *Bot Abh* 4, 1–135.
- Burmester, A., Karimi, S., Wetzel, J. und Wöstemeyer, J. (2013). Complementation of a stable Met2-1 mutant of the zygomycete *Absidia glauca* by the corresponding wild-type allele of the mycoparasite *Parasitella parasitica*, transferred during infection. *Microbiology* 159, 1639–1648.
- Burmester, A., Richter, M., Schultze, K., Voelz, K., Schachtschabel, D., Boland, W., Wöstemeyer, J. und Schimek, C. (2007). Cleavage of β -carotene as the first step in sexual hormone synthesis in zygomycetes is mediated by a trisporic acid regulated β -carotene oxygenase. *Fungal Genet Biol* 44, 1096–1108.
- Burmester, A. und Wöstemeyer, J. (1994). Variability in genome organization of the zygomycete *Parasitella parasitica*. *Curr Genet* 26, 456–460.
- Canutescu, A.A. und Dunbrack, R.L. (2003). Cyclic coordinate descent: a robotics algorithm for protein loop closure. *Protein Sci* 12, 963–972.
- Chen, J.L. und Pace, N.R. (1997). Identification of the universally conserved core of ribonuclease P RNA. *RNA* 3, 557–560.
- Chou, P.Y. und Fasman, G.D. (1978). Empirical predictions of protein conformation. *Ann Rev Biochem* 47, 251–276.
- Consortium, The UniProt (2015). UniProt: a hub for protein information. *Nucleic Acids Res* 43, D204–D212.
- Cornell, M.J., Alam, I., Soanes, D.M., Wong, H.M., Hedeler, C., Paton, N.W., Rattray, M., Hubbard, S.J., Talbot, N.J. und Oliver, S.G. (2007). Comparative genome analysis

- across a kingdom of eukaryotic organisms: Specialization and diversification in the fungi. *Genome Res* 17, 1809–1822.
- Crooks, G.E., Hon, G., Chandonia, J.-M. und Brenner, S.E. (2004). WebLogo: A sequence logo generator. *Genome Res* 14, 1188–1190.
- Czempinski, K., Kruft, V., Wöstemeyer, J. und Burmester, A. (1996). 4-Dihydromethyltrisporate dehydrogenase from *Mucor mucedo*, an enzyme of the sexual hormone pathway: purification, and cloning of the corresponding gene. *Microbiology* 142, 2647–2654.
- Dalby, A., Nourse, J.G., Hounshell, W.D., Gushurst, A.K.I., Grier, D.L., Leland, B.A. und Laufer, J. (1992). Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited. *J Chem Inf Comput Sci* 32, 244–255.
- Dalgaard, J.Z., Garrett, R.A. und Belfort, M. (1994). Purification and characterization of two forms of I-DmoI, a thermophilic site-specific endonuclease encoded by an archaeal intron. *J Biol Chem* 269, 28885–28892.
- Dalgaard, J.Z., Klar, A.J., Moser, M.J., Holley, W. R., Chatterjee, A. und Mian, I.S. (1997). Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family. *Nucleic Acids Res* 25, 4626–4638.
- Danchin, E.G.J., Guzeeva, E.A., Mantelin, S., Berepiki, A. und Jones, J.T. (2016). Horizontal gene transfer from bacteria has enabled the plant-parasitic nematode *Globodera pallida* to feed on host-derived sucrose. *Mol Biol Evol* 33, 1571–1579.
- Dill, K.A. und MacCallum, J.L. (2012). The protein-folding problem, 50 years on. *Science* 338, 1042–1046.
- Duan, Y., Wang, L. und Kollman, P.A. (1998). The early stage of folding of villin head-piece subdomain observed in a 200-nanosecond fully solvated molecular dynamics simulation. *Proc. Natl. Acad. Sci. USA* 95, 9897–9902.
- Dürichen, H., Siegmund, L., Burmester, A., Fischer, M.S. und Wöstemeyer, J. (2016). Ingestion and digestion studies in *Tetrahymena pyriformis* based on chemically modified microparticles. *Eur J Protistol* 52, 45–57.
- Ellenberger, S., Burmester, A., Schuster, S. und Wöstemeyer, J. (2017). Post-translational regulation by structural changes of 4-dihydromethyltrisporate dehydrogenase, a key enzyme in sexual and parasitic communication mediated by the trisporic acid pheromone system, of the fungal fusion parasite *Parasitella parasitica*. *J Theor Biol* 413, 50–57.
- Ellenberger, S., Burmester, A. und Wöstemeyer, J. (2014). Complete mitochondrial DNA sequence of the mucoralean fusion parasite *Parasitella parasitica*. *Genome Announc* 2, e00912–14.

- Ellenberger, S., Burmester, A. und Wöstemeyer, J. (2016a). Complete mitochondrial DNA sequence of the Mucoralean fungus *Absidia glauca*, a model for studying host-parasite interactions. *Genome Announc* 4, e00153–16.
- Ellenberger, S., Burmester, A. und Wöstemeyer, J. (2016b). The fate of mitochondria after infection of the Mucoralean fungus *Absidia glauca* by the fusion parasite *Parasitella parasitica*: comparison of mitochondrial genomes in zygomycetes. *Mitoch DNA Part A*, doi:10.1080/24701394.2016.1248432.
- Ellenberger, S., Schuster, S. und Wöstemeyer, J. (2013). Correlation between sequence, structure and function for trisporoid processing proteins in the model zygomycete *Mucor mucedo*. *J Theor Biol* 320, 66–75.
- Ellenberger, S., Siegmund, L., Schuster, S. und Wöstemeyer, J. (2016). Horizontal gene transfer between bacteria and protozoa: Gene-specific detection by combining different approaches in a new score-based algorithm. *Endocytobiosis Cell Res* 27, 36–46.
- Emameh, R.Z., Barker, H.R., Tolvanen, M.E.E., Parkkila, S. und Hytönen, V.P. (2016). Horizontal transfer of β -carbonic anhydrase genes from prokaryotes to protozoans, insects, and nematodes. *Parasites & Vectors* 9, 152.
- Fávaro, L.C., de Araújo, W.L., de Azevedo, J.L. und Paccola-Meirelles, L.D. (2005). The biology and potential for genetic research of transposable elements in filamentous fungi. *Genet Mol Biol* 28, 804–813.
- Finnegan, D.J. (1989). Eukaryotic transposable elements and genome evolution. *Trends Genet* 5, 103–107.
- Fischer, E. (1894). Einfluss der Configuration auf die Wirkung der Enzyme. *Ber Dtsch Chem Ges* 27, 2985–2993.
- Fitzpatrick, D.A. (2012). Horizontal gene transfer in fungi. *FEMS Microbiol Lett* 329, 1–8.
- Fitzpatrick, D.A., Logue, M.E. und Butler, G. (2008). Evidence of recent interkingdom horizontal gene transfer between bacteria and *Candida parapsilosis*. *BMC Evol Biol* 8, 181.
- Frickey, T. und Lupas, A.N. (2004). PhyloGenie: Automated phylome generation and analysis. *Nucleic Acids Res* 32, 5231–5238.
- Ghosh, D., Sawicki, M., Pletnev, V., Erman, M., Ohno, S., Nakajin, S. und Duax, W. (2001). Porcine carbonyl reductase. Structural basis for a functional monomer in short chain dehydrogenases/reductases. *J Biol Chem* 276, 18457–18463.
- Gilbert, C., Schaack, S., II, J.K. Pace, Brindley, P.J. und Feschotte, C. (2010). A role for host-parasite interactions in the horizontal transfer of transposons across phyla. *Nature* 464, 1347–1350.
- Gladyshev, E.A., Meselson, M. und Arkhipova, I.R. (2008). Massive horizontal gene transfer in Bdelloid rotifers. *Science* 320, 1210–1213.

- Gogarten, J.P. und Townsend, J.P. (2005). Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol* 3, 679–687.
- Görtz, H.-D. (2010). Microbial infections in free-living protozoa. *Crit Rev Immunol* 30, 95–106.
- Grigoriev, I.V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otiillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., Smirnova, T., Nordberg, H., Dubchak, I. und Shabalov, I. (2014). MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res* 42, D699–D704.
- Halary, S., Daubois, L., Terrat, Y., Ellenberger, S., Wöstemeyer, J. und Hijri, M. (2013). Mating type gene homologues and putative sex pheromone-sensing pathway in arbuscular mycorrhizal fungi, a presumably asexual plant root symbiont. *PLoS ONE* 8, e80729.
- Heckmann, K. (1975). Omikron, ein essentieller Endosymbiont von *Euplotes aediculatus*. *J Protozool* 22, 97–104.
- Heinemann, J.A. und Sprague, G.F. (1989). Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. *Nature* 340, 205–209.
- Hoffmann, K., Pawłowska, J., Walther, G., Wrzosek, M., de Hoog, G.S., Benny, G.L., Kirk, P.M. und Voigt, K. (2013). The family structure of the Mucorales: a synoptic revision based on comprehensive multigene-genealogies. *Persoonia* 30, 57–76.
- Hollingsworth, S.A. und Karplus, P.A. (2010). A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *BioMol Concepts* 1, 271–283.
- Hu, W., Walters, B.T., Kan, Z.-Y., Mayne, L., Rosen, L.E., Marqusee, S. und Englander, S.W. (2013). Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. *Proc. Natl. Acad. Sci. USA* 110, 7684–7689.
- Huang, J., Mullapudi, N., Lancto, C.A., Scott, M., Abrahamsen, M.S. und Kissinger, J.C. (2004). Phylogenomic evidence supports past endosymbiosis, intracellular and horizontal gene transfer in *Cryptosporidium parvum*. *Genome Biol* 5, R88.
- Huang, J., Mullapudi, N., Sicheritz-Ponten, T. und Kissinger, J.C. (2004). A first glimpse into the pattern and scale of gene transfer in the Apicomplexa. *Int J Parasitol* 34, 265–274.
- Jefferys, B.R., Kelley, L.A. und Sternberg, M.J.E. (2010). Protein folding requires crowd control in a simulated cell. *J Mol Biol* 397, 1329–1338.
- Jeon, K.W. (2006). *Prokaryotic symbionts of amoeba and flagellates*. In: *The Prokaryotes*. Hrsg. von M. Dworkin, S. Falkow, E. Rosenberg, K.H. Schleifer und E. Stackebrandt. 3rd. 1028-1037. Springer-Verlag New York.
- Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292, 195–202.
- Jones, D.T., Taylor, W.R. und Thornton, J.M. (1992). A new approach to protein fold recognition. *Nature* 358, 86–89.

- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R. und Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240.
- Juhas, M., van der Meer, J.R., Gaillard, M., Harding, R.M., Hood, D.W. und Crook, D.W. (2009). Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol Rev* 33, 376–393.
- Kabsch, W. und Sander, C. (1983). Dictionary of Protein Secondary Structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577–2637.
- Kallberg, Y., Oppermann, U., Jörnvall, H. und Persson, B. (2002). Short-chain dehydrogenases/reductases (SDRs) Coenzyme-based functional assignments in completed genomes. *Eur J Biochem* 269, 4409–4417.
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. und Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 40, D109–D114.
- Kavanagh, K.L., Jörnvall, H., Persson, B. und Oppermann, U. (2008). The SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. *Cell Mol Life Sci* 65, 3895–3906.
- Kavanagh, K.L., Klimacek, M., Nidetzky, B. und Wilson, D.K. (2002). The structure of apo and holo forms of xylose reductase, a dimeric aldo-keto reductase from *Candida tenuis*. *Biochemistry* 41, 8785–8795.
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. und Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols* 10, 845–858.
- Kellner, M., Burmester, A., Wöstemeyer, A. und Wöstemeyer, J. (1993). Transfer of genetic information from the mycoparasite *Parasitella parasitica* to its host *Absidia glauca*. *Curr Genet* 23, 334–337.
- Khor, B.Y., Tye, G.J., Lim, T.S. und Choong, Y.S. (2015). General overview on structure prediction of twilight-zone proteins. *Theor Biol Med Model* 12, 15.
- Kinsey, J.A. (1977). Direct selective procedure for isolating *Neurospora* mutants defective in nicotinamide adenine dinucleotide phosphate-specific glutamate dehydrogenase. *J Bacteriol* 132, 751–756.
- Kinsey, J.A. und Helber, J. (1989). Isolation of a transposable element from *Neurospora crassa*. *Proc. Natl. Acad. Sci. USA* 86, 1929–1933.
- Kirubakaran, P., Pfeiferová, L., Boušová, K., Bednarova, L., Obšilová, V. und Vondrášek, J. (2016). Artificial proteins as allosteric modulators of PDZ3 and SH3 in two-domain constructs: A computational characterization of novel chimeric proteins. *Proteins* 84, 1358–1374.

- Kistler, H.C. und Miao, V.P. (1992). New modes of genetic change in filamentous fungi. *Annu Rev Phytopathol* 30, 131–153.
- Klimacek, M., Wührer, F., Kavanagh, K.L., Wilson, D.K. und Nidetzky, B. (2003). Altering dimer contacts in xylose reductase from *Candida tenuis* by site-directed mutagenesis: structural and functional properties of R180A mutant. *Chem Biol Interact* 143–144, 523–532.
- Koonin, E.V., Makarova, K.S. und Aravind, L. (2001). Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol* 55, 709–742.
- Koonin, E.V., Wolf, Y.I. und Karev, G.P. (2002). The structure of the protein universe and genome evolution. *Nature* 420, 218–223.
- Koski, L.B. und Golding, G.B. (2001). The closest BLAST hit is often not the nearest neighbor. *J Mol Evol* 52, 540–542.
- Kowalski, J.C., Belfort, M., Stapleton, M.A., Holpert, M., Dansereau, J.T., Pietrokovski, S., Baxter, S.M. und Derbyshire, V. (1999). Configuration of the catalytic GIY-YIG domain of intron endonuclease I-Tev I: coincidence of computational and molecular findings. *Nucleic Acids Res* 27, 2115–2125.
- Kozakov, D., Beglov, D., Bohnuud, T., Mottarella, S.E., Xia, B., Hall, D.R. und Vajda, S. (2013). How good is automated protein docking? *Proteins* 81, 2159–2166.
- Kozakov, D., Brenke, R., Comeau, S.R. und Vajda, S. (2006). PIPER: an FFT-based protein docking program with pairwise potentials. *Proteins* 65, 392–406.
- Lang, B.F., Gray, M.W. und Burger, G. (1999). Mitochondrial genome evolution and the origin of eukaryotes. *Annu Rev Genet* 33, 351–397.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. und Higgins, D.G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948.
- Lau, K.F. und Dill, K.A. (1989). A lattice statistical mechanics model of the conformational and sequence spaces of proteins. *Macromolecules* 22, 3986–3997.
- Leaché, A.D. und Rannala, B. (2010). The accuracy of species tree estimation under simulation: A comparison of methods. *Syst Biol* 60, 126–137.
- Lee, J. und Young, J.P.W. (2009). The mitochondrial genome sequence of the arbuscular mycorrhizal fungus *Glomus intraradices* isolate 494 and implications for the phylogenetic placement of *Glomus*. *New Phytol* 183, 200–211.
- Leung, S.-Y., Huang, Y., Lau, S.K.P. und Woo, P.C.Y. (2014). Complete mitochondrial genome sequence of *Lichtheimia ramosa* (syn. *Lichtheimia hongkongensis*). *Genome Announc* 2, e00644–14.
- Levinthal, C. (1968). Are there pathways for protein folding? *J Chim Phys* 65, 44–45.
- Linderstrøm-Lang, K. U. (1952). *Lane medical lectures: proteins and enzymes*. Stanford Univ. Press, Stanford California.

- Lohse, M., Drechsel, O., Kahlau, S. und Bock, R. (2013). OrganellarGenomeDRAW – a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res* 41, W575–W581.
- MacAlpine, D.M., Kolesar, J., Okamoto, K., Butow, R.A. und Perlman, P.S. (2001). Replication and preferential inheritance of hypersuppressive petite mitochondrial DNA. *EMBO J* 20, 1807–1817.
- Madadkar-Sobhani, A. und Guallar, V. (2013). PELE web server: atomistic study of biomolecular systems at your fingertips. *Nucleic Acids Res* 41, W322–W328.
- Mann, M., Smith, C., Rabbath, M., Edwards, M., Will, S. und Backofen, R. (2009). CPSP-web-tools: a server for 3D lattice protein studies. *Bioinformatics* 25, 676–677.
- Marri, P.R., Hao, W. und Golding, G.B. (2007). The role of laterally transferred genes in adaptive evolution. *BMC Evol Biol* 7, S8.
- Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M. und Penny, D. (2002). Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. USA* 99, 12246–12251.
- Matsuo, J., Oguri, S., Nakamura, S., Hanawa, T., Fukumoto, T., Hayashi, Y., Kawaguchi, K., Mizutani, Y., Yao, T., Akizawa, K., Suzuki, H., Simizu, C., Matsuno, K., Kamiya, S. und Yamaguchi, H. (2010). Ciliates rapidly enhance the frequency of conjugation between *Escherichia coli* strains through bacterial accumulation in vesicles. *Res Microbiol* 161, 711–719.
- McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. USA* 36, 344–355.
- Meinke, L., Copeland, A., Lapidus, A., Lucas, S., Berry, K.W., Rio, T. Glavina Del, Hammon, N., Dalin, E., Tice, H., Pitluck, S., Richardson, P., Bruce, D., Goodwin, L., Han, C., Tapia, R., Detter, J.C., Schmutz, J., Brettin, T., Larimer, F., Land, M., Hauser, L., Kyrpides, N.C., Ivanova, N., Göker, M., Woyke, T., Wu, Q.L., Pöckl, M., Hahn, M.W. und Klenk, H.-P. (2012). Complete genome sequence of *Polynucleobacter necessarius* subsp. *asymbioticus* type strain (QLWP1DMWA-1^T). *Stand Genomic Sci* 6, 74–83.
- Miller, M.L. und Sutter, R.P. (1984). Methyl trisporate E. A sex pheromone in *Phycomyces blakesleeanus*. *J Biol Chem* 259, 6420–6422.
- Mizrokhi, L.J. und Mazo, A.M. (1990). Evidence for horizontal transmission of the mobile element *jockey* between distant *Drosophila* species. *Proc. Natl. Acad. Sci. USA* 87, 9216–9220.
- Morris, A.L., MacArthur, M.W., Hutchinson, E.G. und Thornton, J.M. (1992). Stereochemical quality of protein structure coordinates. *Proteins* 12, 345–364.

- Mukherjee, S., Szilagyi, A., Roy, A. und Zhang, Y. (2010). *Genome-wide protein structure prediction*. In: *Multiscale approaches to protein modeling*. Hrsg. von A. Kolinski. Springer-Verlag New York.
- Nagano, K. (1973). Logical analysis of the mechanism of protein folding. *J Mol Biol* 75, 401–420.
- Nakamura, Y., Gojobori, T. und Ikemura, T. (2000). Codon usage tabulated from the international DNA sequence databases: status for the year 2000. *Nucleic Acids Res* 28, 292.
- Nakhleh, L., Ruths, D. und Wang, L.-S. (2005). *RIATA-HGT: A fast and accurate heuristic for reconstructing horizontal gene transfer*. In: *Proceeding COCOON'05 Proceedings of the 11th annual international conference on Computing and Combinatorics*. Hrsg. von L. Wang. Springer-Verlag Berlin, Heidelberg, 84–93.
- Nastri, F., Chino, M., Maglio, O., Bhagi-Damodaran, A., Lu, Y. und Lombardi, A. (2016). Design and engineering of artificial oxygen-activating metalloenzymes. *Chem Soc Rev* 45, 5020–5054.
- Nguyen, M., Ekstrom, A., Li, X. und Yin, Y. (2015). HGT-Finder: A new tool for horizontal gene transfer finding and application to *Aspergillus* genomes. *Toxins* 7, 4035–4053.
- Nocedal, J. und Wright, S.J. (1999). *Numerical optimization*. Hrsg. von P. Glynn und S.M. Robinson. Springer Series in Operations Research. Springer-Verlag Berlin.
- Oppendoes, F.R. und Michels, P.A.M. (2007). Horizontal gene transfer in trypanosomatids. *Trends Parasitol* 23, 470–476.
- Oppermann, U., Filling, C., Hult, M., Shafqat, N., Wu, X., Lindh, M., Shafqat, J., Nordling, E., Kallberg, Y., Persson, B. und Jörnvall, H. (2003). Short-chain dehydrogenases/reductases (SDR): the 2002 update. *Chem Biol Interact* 143–144, 247–253.
- Paquin, B., Laforest, M.J. und Lang, B.F. (1994). Interspecific transfer of mitochondrial genes in fungi and creation of a homologous hybrid gene. *Proc. Natl. Acad. Sci. USA* 91, 11807–11810.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. und Ferrin, T.E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *J Comput Chem* 25, 1605–1612.
- Podell, S. und Gaasterland, T. (2007). Darkhorse: A method for genome-wide prediction of horizontal gene transfer. *Genome Biol* 8, R16.
- Polz, M.F., Alm, E.J. und Hanage, W.P. (2013). Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends Genet* 29, 170–175.
- Popa, O. und Dagan, T. (2011). Trends and barriers to lateral gene transfer in prokaryotes. *Curr Opin Microbiol* 14, 615–623.
- Ragan, M.A. (2001). Detection of lateral gene transfer among microbial genomes. *Curr Opin Genet Dev* 11, 620–626.

- Ravenhall, M., Škunca, N., Lassalle, F. und Dessimoz, C. (2015). Inferring horizontal gene transfer. *PLoS Comput Biol* 11, e1004095.
- Remmert, M., Biegert, A., Hauser, A. und Söding, J. (2012). HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods* 9, 173–175.
- Renner, S.S. und Bellot, S. (2012). *Horizontal gene transfer in eukaryotes: fungi-to-plant and plant-to-plant transfers of organellar DNA*. In: *Genomics of chloroplasts and mitochondria*. Hrsg. von R. Bock und V. Knoop. Springer, Heidelberg.
- Richards, T.A., Soanes, D.M., Foster, P.G., Leonard, G., Thornton, C.R. und Talbot, N.J. (2009). Phylogenomic analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants and fungi. *Plant Cell* 21, 1897–1911.
- Robertson, H.M. (1995). The *Tel-mariner* superfamily of transposons in animals. *J Insect Physiol* 41, 99–105.
- Roper, M., Ellison, C., Taylor, J.W. und Glass, N.L. (2011). Nuclear and genome dynamics in multinucleate ascomycete fungi. *Curr Biol* 21, R786–R793.
- Rose, P.W., Prlic, A., Bi, C., Bluhm, W.F., Christie, C.H., Dutta, S., Green, R.K., Goodsell, D.S., Westbrook, J.D., Woo, J., Young, J., Zardecki, C., Berman, H.M., Bourne, P.E. und Burley, S.K. (2015). The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. *Nucleic Acids Res* 43, D345–D356.
- Rosewich, U.L. und Kistler, H.C. (2000). Role of horizontal gene transfer in the evolution of fungi. *Annu Rev Phytopathol* 38, 325–363.
- Rossmann, M.G., Adams, M.J., Buehner, M., Ford, G.C., Hackert, M.L., Liljas, A. und Rao, S.T. (1973). Molecular symmetry axes and subunit interfaces in certain dehydrogenases. *J Mol Biol* 76, 533–537.
- Rossmann, M.G., Moras, D. und Olsen, K.W. (1974). Chemical and biological evolution of a nucleotide-binding protein. *Nature* 250, 194–199.
- Rost, B. (1999). Twilight zone of protein sequence alignments. *Protein Eng* 12, 85–94.
- Rost, B. und Sander, C. (1993). Prediction of protein secondary structure at better than 70 % accuracy. *J Mol Biol* 232, 584–599.
- Rotkiewicz, P. und Skolnick, J. (2008). Fast procedure for reconstruction of full-atom protein models from reduced representations. *J Comput Chem* 29, 1460–1465.
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A. und Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* 16, 944–945.
- Sander, C. und Schneider, R. (1991). Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins* 9, 56–68.
- Schattner, P., Brooks, A.N. und Lowe, T.M. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res* 33, W686–W689.

- Schimek, C., Petzold, A., Schultze, K., Wetzel, J., Wolschendorf, F., Burmester, A. und Wöstemeyer, J. (2005). 4-Dihydromethyltrisporate dehydrogenase, an enzyme of the sex hormone pathway in *Mucor mucedo*, is constitutively transcribed but its activity is differently regulated in (+) and (-) mating types. *Fung Genet Biol* 42, 804–812.
- Schimek, C. und Wöstemeyer, J. (2006). *Pheromone action in the fungal groups Chytridiomycota, and Zygomycota, and in the Oomycota*. In: *The Mycota I: Growth, Differentiation and Sexuality*. Hrsg. von U. Kües, R. Fischer und K. Esser. Springer-Verlag Berlin, Heidelberg.
- Schimek, C. und Wöstemeyer, J. (2009). Carotene derivatives in sexual communication of zygomycete fungi. *Phytochemistry* 70, 1867–1875.
- Schlegel, M. und Meisterfeld, R. (2003). The species problem in protozoa revisited. *Eur J Protistol* 39, 349–355.
- Schlimme, W., Marchiani, M., Hanselmann, K. und Jenni, B. (1997). Gene transfer between bacteria within digestive vacuoles of protozoa. *FEMS Microbiol Ecol* 23, 239–247.
- Schönknecht, G., Weber, A.P.M. und Lercher, M.J. (2014). Horizontal gene acquisitions by eukaryotes as drivers of adaptive evolution. *BioEssays* 36, 9–20.
- Schultze, K., Schimek, C., Wöstemeyer, J. und Burmester, A. (2005). Sexuality and parasitism share common regulatory pathways in the fungus *Parasitella parasitica*. *Gene* 348, 33–44.
- Schulz, E. und Wetzel, J. (2016). Morphological characterization of sex-deficient mutants of the homothallic zygomycete *Zygorhynchus moelleri*. *Mycoscience* 57, 334–348.
- Schwede, T. (2013). Protein Modeling: What Happened to the „Protein Structure Gap“? *Structure* 21, 1531–1540.
- Seif, E., Leigh, J., Liu, Y., Roewer, I., Forget, L. und Lang, B.F. (2005). Comparative mitochondrial genomics in zygomycetes: bacteria-like RNase P RNAs, mobile elements and a close source of the group I intron invasion in angiosperms. *Nucleic Acids Res* 33, 734–744.
- Seif, E.R., Forget, L., Martin, N.C. und Lang, B.F. (2003). Mitochondrial RNase P RNAs in ascomycete fungi: lineage-specific variations in RNA secondary structure. *RNA* 9, 1073–1083.
- Shakhnovich, E.I. und Gutin, A.M. (1991). Influence of point mutations on protein structure: probability of a neutral mutation. *J Theor Biol* 149, 537–546.
- Sharp, P.M. und Li, W.-H. (1987). The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* 15, 1281–1295.
- Shen, M.-Y. und Freed, K.F. (2002). All-Atom fast protein folding simulations: the villin headpiece. *Proteins* 49, 439–445.

- Sicheritz-Pontén, T. und Andersson, S.G.E. (2001). A phylogenomic approach to microbial evolution. *Nucl Acids Res* 29, 545–552.
- Siegmund, L., Burmester, A., Fischer, M.S. und Wöstemeyer, J. (2013). A model for endosymbiosis: Interaction between *Tetrahymena pyriformis* and *Escherichia coli*. *Eur J Protistol* 49, 552–563.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D. und Higgins, D.G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7, 539.
- Silva, J.C., Loreto, E.L. und Clark, J.B. (2004). Factors that affect the horizontal transfer of transposable elements. *Curr Issues Mol Biol* 6, 57–72.
- Sippl, M.J. (1990). Calculation of conformational ensembles from potentials of mean force. *J Mol Biol* 213, 859–883.
- Sippl, M.J. (1993a). Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures. *J Comput Aided Mol Des* 7, 473–501.
- Sippl, M.J. (1993b). Recognition of errors in three-dimensional structures of proteins. *Proteins* 17, 355–362.
- Sippl, M.J. (1995). Knowledge-based potentials for proteins. *Curr Opin Struct Biol* 5, 229–235.
- Söding, J. (2005). Protein homology detection by HMM-HMM comparison. *Bioinformatics* 21, 951–960.
- Sousa, S.F., Fernandes, P.A. und Ramos, M.J. (2006). Protein-ligand docking: Current status and future challenges. *Proteins* 65, 15–26.
- Stanke, M. und Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res* 33, W465–W467.
- Stoco, P.H., Wagner, G., Talavera-Lopez, C., Gerber, A., Zaha, A., Thompson, C.E., Bartholomeu, D.C., Lückemeyer, D.D., Bahia, D., Loreto, E., Prestes, E.B., Lima, F.M., Rodrigues-Luiz, G., Vallejo, G.A., da Silveira Filho, J.F., Schenkman, S., Monteiro, K.M., Tyler, K.M., de Almeida, L.G.P., Ortiz, M.F., Chiurillo, M.A., de Moraes, M.H., de Lima Cunha, O., Mendonca-Neto, R., Silva, R., Teixeira, S.M.R., I.M.F. Murta, Sincero, T.C.M., de Oliveira Mendes, T.A., Urmenyi, T.P., Silva, V.G., DaRocha, W.D., Andersson, B., Romanha, A.J., Steindel, M., de Vasconcelos, A.T.R. und Grisard, E.C. (2014). Genome of the avirulent human-infective trypanosome - *Trypanosoma rangeli*. *PLoS Negl Trop Dis* 8, e3176.
- Stoddard, B.L. (2006). Homing endonuclease structure and function. *Q Rev Biophys* 38, 49–95.
- Stoddard, B.L. (2014). Homing endonucleases from mobile group I introns: discovery to genome engineering. *Mobile DNA* 5, 7.

- Striepen, B., Pruijssers, A.J.P., Huang, J., Li, C., Gubbels, M.-J., Umejiego, N.N., Hedstrom, L. und Kissinger, J.C. (2004). Gene transfer in the evolution of parasite nucleotide biosynthesis. *Proc. Natl. Acad. Sci. USA* 101, 3154–3159.
- Sui, X., Kiser, P.D., von Lintig, J. und Palczewski, K. (2013). Structural basis of carotenoid cleavage: From bacteria to mammals. *Arch Biochem Biophys* 539, 203–213.
- Sun, B.-F., Li, T., Xiao, J.-H., Jia, L.J., Liu, L., Zhang, P., Murphy, R.W., He, S.-M. und Huang, D.-W. (2015). Horizontal functional gene transfer from bacteria to fishes. *Sci Rep* 5, 18676.
- Surekha, K., Nachiappan, M., Prabhu, D., Choubey, S.K., Biswal, J. und Jeyakanthan, J. (2017). Identification of potential inhibitors for oncogenic target of dihydroorotate dehydrogenase using *in silico* approaches. *J Mol Struct* 1127, 675–688.
- Sutter, R.P., Dadok, J., Bothner-By, A.A., Smith, R.R. und Mishra, P.K. (1989). Cultures of separated mating types of *Blakeslea trispora* make D and E forms of trisporic acids. *Biochemistry* 28, 4060–4066.
- Szöllösi, G.J., Davín, A.A., Tannier, E., Daubin, V. und Boussau, B. (2015). Genome-scale phylogenetic analysis finds extensive gene transfer among fungi. *Phil Trans R Soc B* 370, 20140335.
- Tanaka, M., Bateman, R., Rauh, D., Vaisberg, E., Ramachandani, S., Zhang, C., Hansen, K.C., Burlingame, A.L., Trautman, J.K., Shokat, K.M. und Adams, C.L. (2005). An unbiased cell morphology-based screen for new, biologically active small molecules. *PLoS Biol* 3, 764–776.
- Terzian, C., Ferraz, C., Demaille, J. und Bucheton, A. (2000). Evolution of the Gypsy endogenous retrovirus in the *Drosophila melanogaster* subgroup. *Mol Biol Evol* 17, 908–914.
- Thomas, J. und Pritham, E.J. (2015). Helitrons, the eukaryotic rolling-circle transposable elements. *Microbiol Spectr* 3, MDNA3-0049-2014.
- Trott, O. und Olson, A.J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31, 455–461.
- Van den Ende, H., Werkman, B.A. und van den Briel, M.L. (1972). Trisporic acid synthesis in mated cultures of the fungus *Blakeslea trispora*. *Arch Microbiol* 86, 175–184.
- Vaughn, J.C., Mason, M.T., Sper-Whitis, G.L., Kuhlman, P. und Palmer, J.D. (1995). Fungal origin by horizontal transfer of a plant mitochondrial group I intron in the chimeric *coxI* gene of *Peperomia*. *J Mol Evol* 41, 563–572.
- Webster, J. (1983). *Pilze, Eine Einführung*. Springer-Verlag Berlin Heidelberg.
- Werkman, B.A. (1976). Localization and partial characterization of a sex-specific enzyme in homothallic and heterothallic *Mucorales*. *Arch Microbiol* 109, 209–213.
- Werkman, B.A. und van den Ende, H. (1974). Trisporic acid synthesis in homothallic and heterothallic *Mucorales*. *J Gen Microbiol* 82, 273–278.

- Werner, S., Schroeter, A., Schimek, C., Vlaic, S., Wöstemeyer, J. und Schuster, S. (2012). Model of the synthesis of trisporic acid in Mucorales showing bistability. *IET Syst Biol* 6, 207–214.
- Wetzel, J., Scheibner, O., Burmester, A., Schimek, C. und Wöstemeyer, J. (2009). 4-Dihydrotrisporin-Dehydrogenase, an enzyme of the sex hormone pathway of *Mucor mucedo*: purification, cloning of the corresponding gene, and developmental expression. *Eukaryotic Cell* 8, 88–95.
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P. und Schulman, A.H. (2007). A unified classification system for eukaryotic transposable elements. *Nature Rev Genet* 8, 973–982.
- Wiederstein, M. und Sippl, M.J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35, W407–W410.
- Williamson, D. (2002). The curious history of yeast mitochondrial DNA. *Nat Rev Genet* 3, 475–481.
- Wöstemeyer, A., Teepe, H. und Wöstemeyer, J. (1990). Genetic interactions in somatic inter-mating type hybrids of the zygomycete *Absidia glauca*. *Curr Genet* 17, 163–168.
- Wöstemeyer, J., Siegmund, L. und Ellenberger, S. (2016). Understanding the establishment of endosymbionts in protozoa: Development of a laboratory model for *Tetrahymena pyriformis*. *Endocytobiosis Cell Res* 27, 11–15.
- Wöstemeyer, J., Wöstemeyer, A., Burmester, A. und Czempinski, K. (1995). Relationships between sexual processes and parasitic interactions in the host-pathogen system *Absidia glauca* - *Parasitella parasitica*. *Can J Bot* 73, S243–S250.
- Xi, Z., Wang, Y., Bradley, R.K., Sugumaran, M., Marx, C.J., Rest, J.S. und Davis, C.C. (2013). Massive mitochondrial gene transfer in a parasitic flowering plant clade. *PLoS Genet* 9, e1003265.
- Xia, X. (2007). An improved implementation of codon adaptation index. *Evol Bioinform Online* 3, 53–58.
- Xie, W. und Sahinidis, N.V. (2006). Residue-rotamer-reduction algorithm for the protein side-chain conformation problem. *Bioinformatics* 22, 188–194.
- Yoshiyama, M., Tu, Z., Kainoh, Y., Honda, H., Shono, T. und Kimura, K. (2001). Possible horizontal transfer of a transposable element from host to parasitoid. *Mol Biol Evol* 18, 1952–1958.
- Zerbino, D.R. und Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18, 812–829.
- Zhang, H.-H., Feschotte, C., Han, M.-J. und Zhang, Z. (2014). Recurrent horizontal transfers of *Chaparev* transposons in diverse invertebrate and vertebrate animals. *Genome Biol Evol* 6, 1375–1386.

- Zhu, Q.Y., Kosoy, M. und Dittmar, K. (2014). HGTector: An automated method facilitating genome-wide discovery of putative horizontal gene transfers. *BMC Genomics* 15, 717.
- Župunski, V., Gubenšek, F. und Kordiš, D. (2001). Evolutionary dynamics and evolutionary history in the RTE clade of non-LTR retrotransposons. *Mol Biol Evol* 18, 1849–1863.

6 Anhang

6.1 Zusätzliches Material

Der beigefügte Datenträger enthält diese Dissertation im PDF-Format. Der Ordner „Publikationen“ enthält die im Rahmen dieser Arbeit entstandenen Publikationen. In dem Ordner „mtDNA“ finden sich die Interpretationen der unvollständig oder gar nicht annotierten mtDNAs von Zygomyceten. In dem Ordner „PDB“ liegen die Dateien, die für die TSP-Proteine modellierten Proteinstrukturen und Trajektorien im PDB-Format. Die Strukturdateien beinhalten die Ergebnisse des Protein-Ligand-Dockings in dem Unterordner „Monomere“ und des Protein-Protein-Dockings in dem Unterordner „Dimere“. In dem Ordner „Videos“ finden sich Videos ausgewählter Protein-Modelle, die zu den Abbildungen in dieser Arbeit gehören. Sie veranschaulichen die Dynamik der TSP1-Proteine und die Spezifität der Liganden- beziehungsweise Protein-Bindung. Eine vollständige Sammlung der Videos ist Teil des ZygoViewers und kann aus dem folgenden GitHub-Verzeichnis heruntergeladen werden.

www.github.com/SabrinaEllenberger/ZygoViewer

Der HGT Calculator steht als ausführbares Programm, inklusive einer Anleitung und einem Testdatensatz als Zusatzmaterial, in dem folgenden GitHub-Verzeichnis zu Verfügung.

www.github.com/SabrinaEllenberger/HGT-Calculator

Zu beachten ist, dass die korrekte Ausführbarkeit des HGT Calculators aufgrund seiner Abhängigkeit von externen Datenbanken (UniProt, NCBI) direkt nach einem Update der jeweiligen Datenbank nicht gewährleistet werden kann. In diesem Fall wird eine zeitnahe Anpassung des Programms vorgenommen und eine aktualisierte Programmversion im GitHub-Verzeichnis zur Verfügung gestellt werden.

Das Programm zur ringförmigen Darstellung des GC-Gehalts von mtDNAs, GC-Draw, findet sich in einem dritten GitHub-Verzeichnis.

www.github.com/SabrinaEllenberger/GC-Draw

Der Umfang des zusätzlichen Materials sprengt die Kapazität einer DVD und musste daher auf andere Weise zugänglich gemacht werden.

6.2 Abkürzungsverzeichnis

AI	Alien-Index
ATCC	American Type Culture Collection
BLAST	Basic Local Alignment Search Tool
CAI	Codon Adaptation Index
CR	Konservierte Region
CBS	Centraalbureau voor Schimmelcultures, Utrecht, Niederlande
DAOM	Plant Research Institute, Department of Agriculture (Mycology), Kanada
DSSP	Dictionary of Protein Secondary Structure, Verzeichnis für Sekundärstrukturen in Proteinen
EGT	endosymbiotischer Gentransfer
ENA	European Nucleotide Archive
ERV	Endogene Retroviren
FSU	Fungal Reference Centre, Jena, Deutschland
GFF	General Feature Format
HGT	horizontaler Gentransfer
HMM	Hidden-Markov-Modell
ITS	Internal Transcribed Spacer
JGI	Joint Genome Institute
KEGG	Kyoto Encyclopedia of Genes and Genomes
LINE	Long Interspersed Nuclear Element
LTR	Long Terminal Repeat
NBRC	NITE Biological Resource Center, Department of Biotechnology, National Institute of Technology and Evaluation, Japan
NCBI	National Center for Biotechnology Information, U.S. National Library of Medicine, USA
NRRL	Agricultural Research Service Culture Collection, USA
PDB	Protein Data Bank
PDB-Format	Dateiformat für die Darstellung von Atom-Koordinaten eines Proteins.
PDBQT	Dateiformat ähnlich dem PDB-Format, das zusätzlich zu den Atom-Koordinaten eines Proteins auch Ladungen und Atomtypen darstellt.
PELE	Protein Energy Landscape Exploration
ProSA	Protein Structure Analysis
RCSB	Research Collaboratory for Structural Bioinformatics
SDF	Structured Data File, Dateiformat für die Darstellung von Atom-Koordinaten einer chemischen Verbindung.
TSP1	4-Dihydromethyltrisorat-Dehydrogenase
TSP2	4-Dihydrotrisorin-Dehydrogenase
TSP3	Carotin-Oxygenase
UCSF	University of California, San Francisco

6.3 Abbildungsverzeichnis

1.1	Trisporoidstrukturen	14
1.2	Homothallische Zygosporientwicklung	14
1.3	Heterothallische Zygosporientwicklung	15
1.4	Sporangien von <i>Absidia glauca</i>	15
1.5	Trisporinsäure-Synthese	16
1.6	Zygosporien von <i>Parasitella parasitica</i>	17
1.7	Fusionsparasitismus von <i>Parasitella parasitica</i>	18
1.8	Mechanismus der Transposition in Eukaryoten	19
1.9	Mechanismus der Homing-Endonukleasen in der mtDNA von Zygomyceten	23
1.10	Unterschied zwischen EGT und HGT	25
1.11	Entkommen eines Futterbakterium aus einer Nahrungsvakuole in <i>Tetrahymena pyriformis</i>	26
1.12	Aufbau von Proteinen auf verschiedenen Ebenen	28
1.13	Torsionswinkel in Aminosäureketten	29
1.14	Ramachandran-Plot	30
1.15	Energielandschaft von Proteinen	32
1.16	Künstliches neuronales Netze zur Vorhersage von Proteinstrukturen . . .	35
1.17	HP-Modelle zur Vorhersage von Proteinstrukturen	36
2.1	Genomassemblierung	39
2.2	Normale Modellierung in Phyre2	42
2.3	Intensive Modellierung in Phyre2	44
2.4	Qualitätskontrolle der Proteinvorhersagen	47
2.5	Das PDB-Format	50
2.6	AutoDock Vina in UCSF Chimera	54
2.7	Energieverlauf lokaler Proteinbewegung	56
2.8	Auswahl der Organismen für Suche nach HGT	58
2.9	HGT Downloader	59
2.10	HGT Downloader: Sequenzsuche	60
2.11	HGT Downloader: BLAST-Suche	62
2.12	HGT Downloader: BLAST-Statistik	63
2.13	HGT Downloader: Level1-Bäume	66
2.14	Multi-Level-Analyse der Isocitrat-Dehydrogenase	67
2.15	HGT Calculator: ResultManager	68
2.16	ZygoViewer: Genvorhersage	70
2.17	ZygoViewer: TSP1-Modelle	71
3.1	BLAST-Treffer für Gene aus <i>Absidia glauca</i>	74

3.2	Lineare Darstellung der mtDNAs von Zygomyceten (Teil1)	78
3.3	Lineare Darstellung der mtDNAs von Zygomyceten (Teil2)	79
3.4	Konservierte Regionen in <i>rnpB</i> -Genen aus Zygomyceten	81
3.5	<i>rps3</i> -Gene in Zygomyceten	82
3.6	Ringförmige Darstellung des GC-Gehalts von mtDNA	85
3.7	Klassifizierung transponierbarer Elemente in Eukaryoten	88
3.8	Vergleich der Proteinstrukturen für TSP1, TSP2 und TSP3	90
3.9	Neighbor-Joining-Baum für TSP1-Sequenzen	92
3.10	C-terminale Sequenzmotive in TSP1-Sequenzen	93
3.11	Sequenzalignment der TSP1-Sequenzen vom Typ1	97
3.11	(Fortsetzung)	98
3.12	Strukturvergleich der TSP1-Proteine vom Typ1	99
3.13	Strukturalignment der TSP1-Sequenzen vom Typ1	100
3.14	TSP1-Strukturen für Zygomyceten mit einer TSP1-Sequenz	102
3.15	TSP1-Strukturen für Zygomyceten mit mehr als einer TSP1-Sequenz . . .	104
3.16	TSP1-Strukturen mit flexiblen Strukturbereichen	105
3.17	TSP1-Strukturen für homothallische Zygomyceten	106
3.18	Alignment mit TSP1-ähnlichem Protein aus <i>Rhizophagus irregularis</i> . . .	107
3.19	Struktur eines TSP1-ähnlichen Proteins aus <i>Rhizophagus irregularis</i> . . .	107
3.20	Alignment mit TSP1-ähnlichem Protein aus <i>Ampelomyces quisqualis</i> . . .	108
3.21	Struktur eines TSP1-ähnlichen Proteins aus <i>Ampelomyces quisqualis</i> . . .	108
3.22	Neighbor-Joining-Baum für TSP2-Sequenzen	110
3.23	Sequenzalignment der TSP2-Sequenzen	114
3.24	Strukturvergleich der TSP2-Proteine	115
3.25	Strukturalignment der TSP2-Proteine	116
3.26	TSP2-Strukturen für Zygomyceten mit einer TSP2-Sequenz	117
3.27	TSP2-Strukturen für Zygomyceten mit mehr als einer TSP2-Sequenz . . .	119
3.28	TSP2-Strukturen für homothallische Zygomyceten	120
3.29	Neighbor-Joining-Baum für TSP3-Sequenzen	121
3.30	Sequenzalignment der TSP3-Sequenzen	123
3.30	(Fortsetzung)	124
3.30	(Fortsetzung)	125
3.31	TSP3-Strukturen für einige Zygomyceten	126
3.32	TSP3-Strukturen für homothallische Zygomyceten	127
3.33	TSP3-Strukturen für <i>Absidia</i> - und <i>Parasitella</i> -Stämmen	127
3.34	Substratspezifität von TSP1-Proteinen	133
3.35	Trisporoidbindung bei TSP1-Isomeren	134
3.36	Substratspezifität von TSP2-Proteinen	138
3.37	Trisporoidbindung bei TSP2-Isomeren	139

3.38	Trisporoidbindung bei TSP1- und TSP2-Proteinen	140
3.39	Vergleich von TSP1 und XYL	141
3.40	TSP1-Dimerisierung: Ein Homodimer aus <i>Mucor circinelloides</i>	142
3.41	TSP1-Monomere aus <i>Absidia glauca</i>	143
3.42	TSP1-Dimere aus <i>Absidia glauca</i>	144
3.43	TSP1-Monomere aus <i>Parasitella parasitica</i>	145
3.44	TSP1-Dimere aus <i>Parasitella parasitica</i>	146
3.45	TSP2-Dimere aus <i>Parasitella parasitica</i>	148
3.46	TSP1-TSP2-Dimere aus <i>Parasitella parasitica</i>	149
3.47	TSP2-TSP1-Dimere aus <i>Parasitella parasitica</i>	150
3.48	TSP1-Dimere aus <i>Absidia glauca</i> und <i>Parasitella parasitica</i>	151
3.49	TSP1-Dimere aus <i>Parasitella parasitica</i> und <i>Absidia glauca</i>	152
3.50	Enzymatisch aktive und inaktive TSP1-Isomere in <i>Parasitella parasitica</i> .	153
3.51	TSP1 <i>in silico</i> Mutanten: Aktivierung von TSP1-Homodimeren	154
3.52	HGT-Werte für Isocitrat-Dehydrogenasen aus <i>Leishmania major</i>	159
4.1	Neighbor-Joining-Baum der Endonukleasen aus Zycomyceten	165

6.4 Wissenschaftliche Beiträge

Mitglied des Organisationskomitees der International Study Group for Systems Biology/Young Study Group for Systems Biology (ISGSB/YSGSB) 10/2016 in Jena

6.4.1 Veröffentlichung von Genomen und Software

1. Ellenberger, S., Burmester, A. und Wöstemeyer, J. (2014). Genome sequence of the mucoralean fusion parasite *Parasitella parasitica*. ENA Projektnummer: PRJEB7124.
2. Ellenberger, S., Burmester, A. und Wöstemeyer, J. (2014). *Parasitella parasitica* strain CBS 412.66 mitochondrion, complete genome. GenBank: KM382275.
3. Ellenberger, S., Burmester, A. und Wöstemeyer, J. (2015). Genome sequence of the zygomycete *Absidia glauca*, a model for studying host-parasite interactions and naturally occurring horizontal gene transfer in eukaryotes. ENA Projektnummer: PRJEB13349.
4. Ellenberger, S., Burmester, A. und Wöstemeyer, J. (2015). *Absidia glauca* strain CBS 101.48 mitochondrion, complete genome. GenBank: KU196782.
5. Ellenberger, S. (2016). HGT Calculator: gezielte Identifikation von horizontalem Gentransfer zwischen Bakterien und Protozoen.
GitHub: www.github.com/SabrinaEllenberger/HGT-Calculator.
6. Ellenberger, S. (2016). ZygoViewer: Visualisierung der Zygomycetengenome und TSP1-Proteine von *Absidia glauca* und *Parasitella parasitica*.
GitHub: www.github.com/SabrinaEllenberger/ZygoViewer.
7. Ellenberger, S. (2017). GC-Draw: Ringförmige Darstellung des GC-Gehalts von mitochondrialer DNA.
GitHub: www.github.com/SabrinaEllenberger/GC-Draw.

6.4.2 Publikationen

Der korrespondierende Autor jeder Publikation ist fettgedruckt.

1. **Ellenberger, S.**, Schuster, S. und Wöstemeyer, J. (2013). Correlation between sequence, structure and function for trisporoid processing proteins in the model zygomycete *Mucor mucedo*. *J Theor Biol* 320, 66–75.
2. Halary, S., Daubois, L., Terrat, Y., Ellenberger, S., Wöstemeyer, J. und **Hijri, M.** (2013). Mating type gene homologues and putative sex pheromone-sensing pathway in arbuscular mycorrhizal fungi, a presumably asexual plant root symbiont. *PLoS ONE* 8, e80729.

3. **Ellenberger, S.**, Burmester, A. und Wöstemeyer, J. (2014). Complete mitochondrial DNA sequence of the mucoralean fusion parasite *Parasitella parasitica*. *Genome Announc* 2, e00912-14.
4. **Ellenberger, S.**, Burmester, A. und Wöstemeyer, J. (2016). Complete mitochondrial DNA sequence of the Mucoralean fungus *Absidia glauca*, a model for studying host-parasite interactions. *Genome Announc* 4, e00153-16.
5. **Ellenberger, S.**, Siegmund, L., Schuster, S. und Wöstemeyer, J. (2016). Horizontal gene transfer between bacteria and protozoa: Gene-specific detection by combining different approaches in a new score-based algorithm. *Endocytobiosis Cell Res* 27, 36–46.
6. **Wöstemeyer, J.**, Siegmund, L. und Ellenberger, S. (2016). Understanding the establishment of endosymbionts in protozoa: Development of a laboratory model for *Tetrahymena pyriformis*. *Endocytobiosis Cell Res* 27, 11–15.
7. **Wöstemeyer, J.**, Ellenberger, S., Schulz, E., Siegmund, L., van Burgeler, A., Gerlitz, N., Burmester, A., Wetzel, J. und Voigt, J. (2016). Fusion parasitism between *Parasitella parasitica* and its host *Absidia glauca*: A system between sexuality and parasitism. *Endocytobiosis Cell Res* 27, 24–32.
8. **Wöstemeyer, J.**, Schimek, C., Wetzel, J., Burmester, A., Voigt, J., Schulz, E., Ellenberger, S. und Siegmund, L. (2016). Wendland, J. (Ed.). Pheromone action in the fungal groups Chytridiomycetes and Zygomycetes and in the Oophytes. *Springer International Publishing Switzerland*.
9. **Ellenberger, S.**, Burmester, A. und Wöstemeyer, J. (2016). The fate of mitochondria after infection of the Mucoralean fungus *Absidia glauca* by the fusion parasite *Parasitella parasitica*: comparison of mitochondrial genomes in zygomycetes. *Mitochondr DNA Part A* doi:10.1080/24701394.2016.1248432.
10. **Schulz, E.**, Wetzel, J., Burmester, A., Ellenberger, S., Siegmund, L. und Wöstemeyer, J. (2016). Sex loci of homothallic and heterothallic Mucorales. *Endocytobiosis Cell Res* 27, 39–57.
11. **Ellenberger, S.**, Burmester, A., Schuster, S. und Wöstemeyer, J. (2017). Post-translational regulation by structural changes of 4-dihydromethyltrisporate dehydrogenase, a key enzyme in sexual and parasitic communication mediated by the trisporic acid pheromone system, of the fungal fusion parasite *Parasitella parasitica*. *J Theor Biol* 413, 50–57.

12. **Ellenberger, S.**, Burmester, A., van Burgeler, A., Schuster, S. und Wöstemeyer, J. (in Vorbereitung). Post-translational regulation in the trisporic acid pheromone system of the fungal fusion parasite *Parasitella parasitica* and its host *Absidia glauca*. *J Biomol Struct Dyn*.

6.4.3 Poster

1. Ellenberger, S., Schuster, S. und Wöstemeyer, J. (2015). Modelling for activity prediction of *in silico* mutants exemplified by trisporic acid synthesis. P03 bei der Central German Meeting on Bioinformatics (CGMB) in Halle.
2. Ellenberger, S., Schuster, S. und Wöstemeyer, J. (2012). Theoretical study of two minus mating type specific dehydrogenases of the zygomycete *Mucor mucedo*. P47 bei der German Conference on Bioinformatics (GCB) in Jena.
3. Ellenberger, S., Schuster, S. und Wöstemeyer, J. (2013). The HGT Calculator: targeted detection of horizontal gene transfer from prokaryotes to protozoa in small data sets. P27 bei der German Conference on Bioinformatics (GCB) in Göttingen.
4. Ellenberger, S., und Wöstemeyer, J. (2013). Structure modeling of proteins for the biosynthesis of sex pheromonones in zygomycetous fungi. P39 bei der German Conference on Bioinformatics (GCB) in Göttingen.
5. Ellenberger, S., Siegmund, L. und Wöstemeyer, J. (2014). A new task for HGT Calculator: Exploring relations between lifestyle and the frequency of horizontal gene transfer in protozoa. P12 bei der German Conference on Bioinformatics (GCB) in Bielefeld.
6. Ellenberger, S., Burmester, A. und Wöstemeyer, J. (2014). The mtDNA of the mycoparasitic fusion parasit *Parasitella parasitica*: Sequence and comparative analysis. P39 bei der German Conference on Bioinformatics (GCB) in Bielefeld.
7. Ellenberger, S., Schuster, S. und Wöstemeyer, J. (2014). Structure modelling, protein-ligand, and protein-protein interaction of sex pheromone processing dehydrogenases in *Mucor*-like fungi. P82 bei der German Conference on Bioinformatics (GCB) in Bielefeld.
8. Ellenberger, S., Schuster, S. und Wöstemeyer, J. (2015). Prediction of interaction-dependent enzyme activity in the trisporic acid pheromone system by modelling of *in silico* mutants. P06 bei der German Conference on Bioinformatics (GCB) in Dortmund.

9. Ellenberger, S., Burmester, A., van Burgeler, A., Schuster, S., Wöstemeyer, J. (2016). Post-translational regulation of enzyme activity investigated by protein-protein docking: Sexual and parasitic communication of the fungal fusion parasite *Parasitella parasitica*. P14 bei der Tagung der International Study Group on Systems Biology (ISGSB) in Jena.
10. Wöstemeyer, J., Burmester, A., Ellenberger, S., Wetzel, J., Voigt, J., Siegmund, L. und Schulz, E. (2015). Parasexual gene transfer from the mycoparasitic fungus *Parasitella parasitica* to its host *Absidia glauca* – Characterization of transferred DNA and sequence analysis of the *Parasitella parasitica* genome. MEvP02 bei der Jahrestagung der Vereinigung für Allgemeine und Angewandte Mikrobiologie (VAAM) in Marburg.
11. Ellenberger, S., Burmester, A. und Wöstemeyer, J. (2016). Complete genomes of the mucoralean fusion parasite *Parasitella parasitica* and its host *Absidia glauca* substantiate the frequency of gene transfer between host and parasite. FBP12 bei der Jahrestagung der Vereinigung für Allgemeine und Angewandte Mikrobiologie (VAAM) in Jena.

6.5 Ehrenwörtliche Erklärung

Ich erkläre hiermit, dass mir die geltende Promotionsordnung der Biologisch-Pharmazeutischen Fakultät der Friedrich-Schiller-Universität Jena bekannt ist,

ich die Dissertation selbst angefertigt habe und alle von mir benutzten Hilfsmittel, persönlichen Mitteilungen und Quellen in meiner Arbeit angegeben sind,

mich folgende Personen bei der Auswahl und Auswertung des Materials, sowie bei der Herstellung des Manuskripts unterstützt haben:

1. Prof. Dr. Johannes Wöstemeyer
2. Prof. Dr. Stefan Schuster
3. PD Dr. Anke Burmester

die Hilfe eines Promotionsberaters nicht in Anspruch genommen wurde und dass Dritte weder unmittelbar noch mittelbar geldwerte Leistungen von mir für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen,

dass ich die Dissertation noch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht habe und

dass ich die gleiche, eine in wesentlichen Teilen ähnliche oder eine andere Abhandlung nicht bei einer anderen Hochschule als Dissertation eingereicht habe.

Jena, 2. Februar 2017

Sabrina Ellenberger